噪声二元分类器下纠错输出编码的可靠性函数

Gendo Kumoi *kumoi@vos.nagaokaut.ac.jp*
长冈技术科学大学信息与管理系统工程系

Hideki Yagi *h.yagi@uec.ac.jp*
电气通信大学计算机与网络工程系

Manabu Kobayashi *mkoba@waseda.jp*
早稻田大学数据科学中心

Shigeichi Hirasawa *hira@waseda.jp*
早稻田大学数据科学中心

arXiv:2611.33068v1 [cs.IT] 24 Nov 2026

摘要

纠错输出编码（ECOC）利用二元分类器构建多分类器。我们研究当每个二元分类器输出的后验估计的对数优势比受到高斯噪声干扰时，ECOC 相对于贝叶斯（MAP）分类器的超额错误率。我们将解码重新表述为高斯信道上的检验，利用并集界和 Chernoff 界推导上界，并通过与最近的竞争类别进行比较以及 Mills 比推导下界；这分别对应于信道编码中的可达性界和逆定理界。界限表明，较大的汉明距离是降低超额错误率的必要且充分条件。只要两个最可能类别的码字之间的距离有界，超额错误率就会与零保持一定距离；这个下界常数仅取决于最大的两个后验概率和噪声水平，而与类别数量无关 M。对于等距码，该距离为 $d_{\min}$。反过来，超额错误率至多为 $M-1$ 乘以一个关于最小距离 $d_{\min}$呈指数级衰减的因子；当 $\kappa\, d_{\min} - \ln(M-1) \to \infty$ 时，它趋于零，其中常数 κ可显式给出（对于一般码，还需满足关于后验概率的条件）。对于 $n \geq 2$的 n-对其余类别编码（一类等距码），上下界的指数渐近一致，因此可以用闭式确定 ECOC 的可靠性函数。将汉明距离替换为有效距离后，结论可推广到相关高斯噪声；当噪声在类别层面共享时，有效距离保持有界，因此对于任何码字表，超额错误率都不会趋于零。在严格满足假设的半合成设置中，测得的衰减与预测指数一致。在 CIFAR-10/100 和 ImageNet-1k 的 ResNet-50 特征上，相对于对数似然解码基准的超额错误随 $d_{\min}$ 下降；而对于类别数较多的情况，则会趋于一个适中的

距离，而当 M 增大时，具有 $d_{\min} = 1$ 的码会退化。

1 引言

多分类问题（$M \geq 3$ 个类别）可以通过组合多个二元分类器来解决，这种方法称为纠错输出编码（ECOC）（Dietterich & Bakiri，1995）。ECOC 因其性能和易于实现而得到广泛应用，而基于损失的解码将其与多分类转二分类问题的其他约简方法统一起来（

Allwein 等人，2000）。其性能取决于所组合的二元分类器；许多数据集上的实证研究已证明了一些有效的组合方式（Dietterich & Bakiri，1995；Rifkin & Klautau，2004；Escalera 等人，2010；Rocha & Goldenstein，2014），但这些组合强烈依赖于数据集，目前尚未建立指导此类选择的一般原则。

在无噪声假设下，即每个二元分类器都能估计真实后验概率，已有研究给出了最优组合的条件（James & Hastie, 1998; Rifkin & Klautau, 2004; Kumoi等人, 2021）。Kumoi等人（2022；2023）提出了一个用于解析评估噪声二元分类器的框架，并展示了汉明距离对成对错误概率——即两个类别之间的错误率——的影响。然而，这一结果仅限于成对错误概率，并未扩展到所有 M 类别上的分类错误率，因此汉明距离增大如何降低ECOC的分类错误率，以及这种作用是否既充分又必要，仍有待探究。

本文针对ECOC的超额错误率 回答了这些问题。超额错误率是指其错误率超过贝叶斯（MAP）分类器错误率的部分（即0–1损失下的超额风险；Bartlett等人, 2006; Tewari & Bartlett, 2007）。本文的贡献如下。(i) 我们将对数优势比噪声下的ECOC解码重新表述为高斯信道上的检验，并将超额错误率化简为ECOC未能选出MAP类别的概率（第3节）。(ii) 我们推导出一个上界（第 4 节）和一个下界（第5节），由此证明，对于每个码字表，较大的最小距离是充分条件；而两个最可能类别对应码字之间较大的距离则是必要条件，且后者对 M一致成立。(iii) 对于 nvR码（$n \geq 2$），这类码是等距码，其两个指数渐近趋于一致，从而以闭式确定了ECOC的可靠性函数（第6节）。(iv) 我们通过有效距离将结果推广到任意高斯噪声协方差；在此条件下，共模相关性无害，而类别级噪声会限制可实现的增益（第8节）。(v) 半合成实验和真实图像实验验证了预测的指数，以及超额错误率随最小距离增大而衰减的趋势（第9节）；我们还指出了限制真实数据上定量吻合度的因素（第10节）。第2–8节中的所有引理、命题、推论和定理，以及附录中的证明，均已在Lean 4和Mathlib中进行机器验证，涵盖从高斯噪声模型和解码器到可靠性函数的内容（本arXiv投稿的附属文件）；数值和实验结果不属于形式化验证的范围。

相关工作。 约简的遗憾分析通过二元问题的超额误差来界定多分类器的超额误差，并且对任意分布均成立。对于采用汉明解码的 ECOC，多类训练误差至多为二元错误率平均值的 $2N/d_{\min}$ 倍（Allwein 等人，2000，推论 2）；误差限制约简则将多类误差界定为二元误差的常数倍（Beygelzimer 等人，2005）；改进约简则将二元遗憾 转化为多类遗憾，但需要付出平方根因子（Langford & Beygelzimer，2005），或者仅带来常数倍增长，并使每个样例的计算量增加O（log M）（Beygelzimer 等人，2009）；Ramaswamy 等人（2014）刻画了采用替代损失的输出编码方法何时具有一致性。在这些最坏情形保证中，最小距离仅以常数因子的形式出现；而在噪声模型下，它决定指数衰减率。Nguyen 等人（2021）在假设二元判决错误相互独立的情况下，证明了汉明解码的错误率随码长呈指数衰减。我们则对二元分类器输出的后验估计中的噪声进行建模，采用似然解码，并从两侧界定相对于 MAP 分类器的超额误差，从而确定指数及其对后验分布的依赖关系。对于随机生成的码，Kobayashi 等人（2024）将 ECOC 与 MAP 估计不一致的平均概率界定为形式为 N（$c - R$）、码率为 $R = \ln M/N$ 的指数；我们的结果针对固定的（构造性）码字表，而充分性条件 $\kappa\, d_{\min} - \ln(M-1) \to \infty$ 是其对应结果。仅使用 $\lceil \log_2 M \rceil$ 位的输出编码也被用于缩减大词汇表神经序列模型的输出层；研究发现普通二元编码并不充分，而误差纠正码带来的冗余能够提高准确率（Oda 等，2017）。我们的必要性结果与这一观察一致。

2 问题设定

2.1 ECOC 与码字表

分类问题是根据给定的输入数据 $\boldsymbol{x} \in \mathbb{R}^D$，估计与之对应的类别 c_i（$i \in \mathcal{C} \triangleq \{1、2、\ldots\ldots、M\}$），其中 $M \geq 3$。在 ECOC 中，类别是根据按照预定义码字表构造的二元分类器的输出进行估计的（Dietterich & Bakiri，1995；Rifkin & Klautau，2004；Escalera 等人，2010）。码字表是一个 $M \times N$ 矩阵 W，其元素属于 $w_{ij} \in \{0, 1\}$；每一行是一个类别的码字，每一列定义一个二元分类器。我们将由 W 的第 j 列构造的二元分类器记为 f_j，并令 $\mathcal{F} \triangleq \{1、2、\ldots\ldots、N\}$ 表示二元分类器集合。

我们假设 W 具有以下性质。前两项不失一般性，因为常数列不会定义分类器，而对一列取补集只会交换分类器的两侧标签；后两项则是必要条件。

- 不存在所有元素均为 0 或均为 1 的列：$0 < \sum_{i \in \mathcal{C}} w_{ij} < M, \quad \forall j \in \mathcal{F}$。
- 每列中 1 的数量至多为 $M/2$：$\sum_{i \in \mathcal{C}} w_{ij} \leq M/2$。
- 所有类别的码字两两不同。
- 所有二元分类器均不相同。

类别 c_i 与 c_k 之间的汉明距离为

$$d_H(c_i, c_k) = \sum_{j=1}^{N} w_{ij} \oplus w_{kj}, \tag{1}$$

其中 $\oplus$ 表示异或；不同的码字对于 $i \neq k$给出 $d_H(c_i, c_k) \geq 1$。我们用 $d_{\min} \triangleq \min_{i \neq k} d_H(c_i, c_k)$ 表示该表的最小距离；在第 2.2 节确定 MAP类别 c_1 后，$d_{\min}(c_1) \triangleq \min_{k \neq 1} d_H(c_1, c_k) \geq d_{\min}$ 表示其码字到最近其他码字的距离；对于等距码，二者相同。

定义 1 (n**vR 码)。** nvR 码是一个码字表，其中包含将 M 个类别划分为 n 与（$M-n$）、$0 < n \leq \lfloor M/2 \rfloor$ 的所有组合；对于 $n = M/2$，每种划分均包含一次。nvR 码的汉明距离为

$$d(n, M) = \begin{cases} 2\binom{M-2}{n-1}, & n \neq \frac{M}{2}, \\ \binom{M-2}{n-1}, & n = \frac{M}{2}. \end{cases} \tag{2}$$

特别地，nvR编码是等距码：对于所有 $i \neq k$，$d_H(c_i, c_k) = d(n, M)$.

定义 2（穷举码）。穷举码是通过串接所有 nvR编码（$1 \leq n \leq \lfloor M/2 \rfloor$）得到的码字表。其码长为 $N_{\text{exh}} = 2^{M-1} - 1$，最小汉明距离为$d_{\min} = 2^{M-2}$。

定义 3（哈达玛码）。对于满足 Hadamard 阶数为 M 的矩阵存在的 $M \equiv 0 \pmod 4$（据推测，每个这样的 M都成立），哈达玛码由阶数为M的归一化Hadamard矩阵通过映射 $+1 \mapsto 0$ 和 $-1 \mapsto 1$得到；其 M 个码字的码长为 M，两两距离为 $M/2$（Levenshtein，1961；MacWilliams与Sloane，1977）。其第一列为常数，因此作为码字表，我们将其删除，留下一个具有 $N = M - 1$ 和 $d = M/2$的等距表，达到 Plotkin界（Plotkin，1960）；它通常是非线性的，而对于具有 Sylvester矩阵的 $M = 2^m$，它是线性单纯形码。Hirasawa等人（2022）、Kumoi等人（2022）和 Hirasawa等人（2022）利用 Reed–Muller码和Hadamard矩阵为 ECOC 构造了这类等距码；他们将码长为 $M - 1$ 的码称为单纯形码。对于 $M = 4$，它与 2vR编码至多相差列取补（第 7.1 节）。

logM 码的码字是类别索引的二进制表示（$N = \lceil \log_2 M \rceil, d_{\min} = 1$），它是最短的表，并作为第 9 节中的小距离极限。

2.2 噪声模型与成对误差概率

我们将输入数据 $\boldsymbol{x}$ 下每个类别的真实后验概率定义为 $\theta_i = P(c_i|\boldsymbol{x})$，并记为$\boldsymbol{\theta} = (\theta_1、\ldots\ldots、\theta_M)$。下文中，$f_j(y|\boldsymbol{\theta})$表示给定 $\boldsymbol{x}$时二元类别 $y \in \{0、1\}$的条件概率。无噪声二元分类器 f_j^* 的输出为

$$f_j^*(1|\boldsymbol{\theta}) = \sum_{i' \in \mathcal{C}: w_{i'j}=1} \theta_{i'}. \tag{3}$$

我们采用基于二元分类器输出的后验概率乘积的解码方法，

$$g(c_i|\boldsymbol{\theta}) = \prod_{j=1}^{N} f_j(w_{ij}|\boldsymbol{\theta}), \qquad \hat{i} = \arg\max_{i \in \mathcal{C}} g(c_i|\boldsymbol{\theta}), \tag{4}$$

并在无噪声情况下记作 $g^*(c_i|\boldsymbol{\theta}) = \prod_{j=1}^{N} f_j^*(w_{ij}|\boldsymbol{\theta})$。

定义 4. 对数后验概率比的误差项 ε_j 为

$$\varepsilon_j = \ln \frac{f_j(1|\boldsymbol{\theta})}{1 - f_j(1|\boldsymbol{\theta})} - \ln \frac{f_j^*(1|\boldsymbol{\theta})}{1 - f_j^*(1|\boldsymbol{\theta})}. \tag{5}$$

若 $f_j^*(1|\boldsymbol{\theta}) \in \{0, 1\}$，则定义 $f_j(1|\boldsymbol{\theta}) = f_j^*(1|\boldsymbol{\theta})$ 和 $\varepsilon_j = 0$。

等价地，

$$f_j(1|\boldsymbol{\theta}) = \frac{e^{\varepsilon_j} f_j^*(1|\boldsymbol{\theta})}{1 + (e^{\varepsilon_j} - 1) f_j^*(1|\boldsymbol{\theta})}. \tag{6}$$

假设1。 $\varepsilon_j \sim \mathcal{N}(0,\ \sigma^2)$，独立同分布

根据（4）和（6）， c_i 和 c_k 的解码分数之比是随机变量 le

$$R_{ik} = \frac{g(c_i|\boldsymbol{\theta})}{g(c_k|\boldsymbol{\theta})} = e^{\sum_{j=1}^{N}(w_{ij} - w_{kj})\varepsilon_j} \frac{g^*(c_i|\boldsymbol{\theta})}{g^*(c_k|\boldsymbol{\theta})}. \tag{7}$$

命题 1（成对错误概率（Kumoi 等，2023））。在假设 1,

$$P(R_{ik} \le 1|\boldsymbol{\theta}) = \Phi\left(\frac{\ln \frac{g^*(c_k|\boldsymbol{\theta})}{g^*(c_i|\boldsymbol{\theta})}}{\sigma \sqrt{d_H(c_i, c_k)}} \right), \tag{8}$$

其中 $\Phi(\cdot)$是标准正态分布的累积分布函数。

我们定义

$$\mu_{ik} \triangleq \ln \frac{g^*(c_k|\boldsymbol{\theta})}{g^*(c_i|\boldsymbol{\theta})}, \qquad \rho_{ik} \triangleq \frac{-\mu_{ik}}{\sigma \sqrt{d_H(c_i, c_k)}}. \tag{9}$$

此后，不失一般性地将类别排序，使得 $\theta_1 > \theta_2 \ge \cdots \ge \theta_M$，并假设$\theta_M > 0$；我们在约束集上进行研究

$$\Omega(\lambda) = \{\boldsymbol{\theta} : \theta_1 \ge \max_{k \ne 1} \theta_k + \lambda\} \tag{10}$$

对于 $\lambda > 0$，其 MAP类别为 c_1。由命题 1 可知

$$P(R_{1k} \le 1|\boldsymbol{\theta}) = \Phi(-\rho_{1k}) = Q(\rho_{1k}), \tag{11}$$

其中 $Q(x) = 1 - \Phi(x)$；当 $\mu_{1k} \le 0$ 时为 $\rho_{1k} \ge 0$，这对满足 $\boldsymbol{\theta} \in \Omega(\lambda)$ 的 nvR 码成立（命题4）。除非另有说明，第3–7节均采用假设1。

3 等效信道与超额误差率

3.1 等效高斯信道

由（6）可得,

$$\ln g(c_i|\boldsymbol{\theta}) = \sum_{j=1}^{N} \Big[\ln f_j^*(w_{ij}|\boldsymbol{\theta}) + w_{ij}\varepsilon_j - \ln\big(1 + (e^{\varepsilon_j} - 1) f_j^*(1|\boldsymbol{\theta})\big) \Big]. \tag{12}$$

第三项 不依赖于 i，因此 ECOC 的决策仅取决于充分统计量 c

$$L_i \triangleq \sum_{j=1}^{N} w_{ij}\varepsilon_j + \ln g^*(c_i|\boldsymbol{\theta}), \qquad \hat{i} = \arg\max_{i \in \mathcal{C}} L_i. \tag{13}$$

相对于 MAP 类别 c_1,

$$L_k - L_1 = \mu_{1k} + Z_{1k}, \qquad Z_{1k} \triangleq \sum_{j=1}^{N} (w_{kj} - w_{1j})\varepsilon_j \sim \mathcal{N}(0, d_H(c_1, c_k)\sigma^2), \quad k \neq 1. \tag{14}$$

这是高斯信道上的一个 M 元假设检验：传输字是 c_1 的码字，竞争类别 c_k 与其之间由均值偏移（信号）μ_{1k} 和距离 $d_H(c_1, c_k)$隔开，而式 (9) 中的 ρ_{1k} 是相应的信噪比。

备注 1（与信道编码的对应关系）。 不同于标准 AWGN信道，信号 μ_{1k} 和噪声方差通过汉明距离依赖于编码字对。对于等距码，这种依赖消失，并且每个成对检验 $L_k - L_1$ 都是标准的二元AWGN判决。$\{Z_{1k}\}_{k\neq 1}$ 仍然具有联合依赖性，因为它们由同一个$\{\varepsilon_j\}$构成，因此该 M元问题并不是一组相互独立的并行信道。

3.2 超额错误率

我们将 ECOC 选择除 c_1 之外的某个类别这一事件定义为

$$E = \bigcup_{k \neq 1} \{R_{1k} \leq 1\}, \tag{15}$$

因此， $P(E|\boldsymbol{\theta})$ 是 ECOC 未能选择 MAP类别的概率。 E 还包含平局情形$R_{1k} = 1$；在假设1（以及第8节的相关噪声）下，每个 ln R_{1k}都是非退化高斯分布，因此平局的概率为零，并且 $P(E|\boldsymbol{\theta}) = P(\hat{i}_{\mathrm{ECOC}} \neq 1|\boldsymbol{\theta})$。我们将在下面的（16）式以及命题 2 的证明中使用这一恒等式。令真实类别为随机变量 C（其中 $P(C = c_i|\boldsymbol{\theta}) = \theta_i$）；给定 $\boldsymbol{\theta}$ 后，噪声$\{\varepsilon_j\}$ 与 C 独立。MAP错误率为 $P_{\mathrm{MAP}}(\boldsymbol{\theta}) = 1 - \theta_1$，而 ECOC 的错误率分解为

$$P_{\mathrm{ECOC}}(\boldsymbol{\theta}) = \theta_1 P(E|\boldsymbol{\theta}) + \sum_{k \neq 1} \theta_k P(\hat{i}_{\mathrm{ECOC}} \neq k \mid \boldsymbol{\theta}). \tag{16}$$

定义 5（超额错误率）。 ECOC 在 $\boldsymbol{\theta}$ 处的（条件）超额错误率为

$$\Delta(\boldsymbol{\theta}) = P_{\mathrm{ECOC}}(\boldsymbol{\theta}) - P_{\mathrm{MAP}}(\boldsymbol{\theta}). \tag{17}$$

由于 P_{MAP} 是 $\boldsymbol{x}$ 处的贝叶斯错误率， $\Delta(\boldsymbol{\theta})$ 是 ECOC 在 0–1损失下的条件超额风险；对输入分布进行平均即可得到通常意义上的超额风险（Bartlett 等，2006；Tewari 和 Bartlett，2007）。将 $1 - \theta_1 = \sum_{k\neq 1} \theta_k$ 代入（16）可得精确分解

$$\Delta(\boldsymbol{\theta}) = \theta_1 P(E|\boldsymbol{\theta}) - \sum_{k \neq 1} \theta_k P(\hat{i}_{\mathrm{ECOC}} = k|\boldsymbol{\theta}). \tag{18}$$

命题 2（超额错误率的双侧评估）。 对于任意码字表以及任意 $\boldsymbol{\theta} \in \Omega(\lambda)$，在任何平局概率为零的噪声下（特别是在假设1以及第8节所述噪声下），

$$(\theta_1 - \theta_2)\, P(E|\boldsymbol{\theta}) \le \Delta(\boldsymbol{\theta}) \le \theta_1 P(E|\boldsymbol{\theta}). \tag{19}$$

特别地，由于 $\theta_1 - \theta_2 \ge \lambda$，我们有 $\Delta(\boldsymbol{\theta}) \ge \lambda\, P(E|\boldsymbol{\theta})$。

证明。上界。 (18) 的第二项非负，因此 $\Delta(\boldsymbol{\theta}) \le \theta_1 P(E|\boldsymbol{\theta})$。

下界。 事件 $\{\hat{i}_{\mathrm{ECOC}} = k\}$ $(k \ne 1)$ 互不相交，且 $\{\hat{i}_{\mathrm{ECOC}} = k\} \subseteq \{R_{1k} \le 1\} \subseteq E$，因此

$$\sum_{k \ne 1} P(\hat{i}_{\mathrm{ECOC}} = k|\boldsymbol{\theta}) = P(\hat{i}_{\mathrm{ECOC}} \ne 1|\boldsymbol{\theta}) \le P(E|\boldsymbol{\theta}). \tag{20}$$

结合 $\theta_k \le \theta_2$ $(k \ne 1)$，

$$\sum_{k \ne 1} \theta_k P(\hat{i}_{\mathrm{ECOC}} = k|\boldsymbol{\theta}) \le \theta_2\, P(E|\boldsymbol{\theta}). \tag{21}$$

将其代入 (18) 可得下界。 □

因此 $\Delta(\boldsymbol{\theta})$ 与 $P(E|\boldsymbol{\theta})$ 相差至多 $[\lambda, \theta_1]$ 中的一个因子，且它们的归一化指数相差至多 $(-\ln \lambda)/d_{\min}(c_1)$；第 4–6 节给出了 $P(E|\boldsymbol{\theta})$ 的界。

3.3 目标：可靠性函数

在信道编码中，可靠性函数是指数 E，错误率随码长 N 按 $\exp(-N \cdot E)$ 衰减。这里的渐近变量是最小距离，而不是 N。根据式（14），统计量 $L_k - L_1$ 只取决于 c_1 与 c_k 不同的位置。因此，信号 μ_{1k} 和噪声方差 $d_H(c_1, c_k)\,\sigma^2$ 都随 d_H 缩放，而一对编码的有效码长是 d_H，而不是 N。此时，误差主要由最近竞争者决定，因此 $P(E|\boldsymbol{\theta})$ 的指数与 $d_{\min}(c_1)$ 成正比，正如编码理论中最小距离决定纠错能力一样。由于 N 和 $d_{\min}$ 在不同码族之间差异很大（对于 2vR，$N = O(M^2)$，但 $d_{\min} = O(M)$；对于 logM 码，$N = O(\log M)$，但 $d_{\min} = 1$；见表 1），最小距离是自然的变量。图 1(a,b) 展示了这一机制。我们的目标是求出确切的指数 E_{ECOC}，使得 $P(E|\boldsymbol{\theta}) \asymp \exp(-d_{\min}(c_1) \cdot E_{\mathrm{ECOC}})$；对于等距码，我们用 d 表示共同距离。

4 上界：最小距离的充分性

策略。 我们对 μ_{1k} 给出上界（命题 3 和 4），从而对 ρ_{1k} 给出下界（例如 $\rho_{1k} \ge \rho^{\mathrm{L}}$，式 (32)）；由于 Q 单调递减，并集界给出 $P(E|\boldsymbol{\theta}) \le (M-1)Q(\rho^{\mathrm{L}})$，而命题 2 将其转化为超额错误率的上界（命题 5、推论 1）。

4.1 μ_{1k}的上界

我们引入

$$\omega(n) \triangleq \begin{cases} 0, & n = 1, \\ \sum_{k=M-(n-2)}^{M} \theta_k, & n > 1, \end{cases} \qquad n_{\min} \triangleq \min_{j \in \mathcal{F}: w_{1j} = 1} \sum_{i \in \mathcal{C}} w_{ij}, \tag{22}$$

其中，$\omega(n)$ 是最小的 $n-1$ 个后验概率之和，$n_{\min}$ 是在其正侧包含 c_1 的各列中，1 的最小数量，并且

$$\alpha_j \triangleq \sum_{i' \in \{\mathcal{C} \setminus \{1\}: w_{1j} = w_{i'j}\}} \theta_{i'}. \tag{23}$$

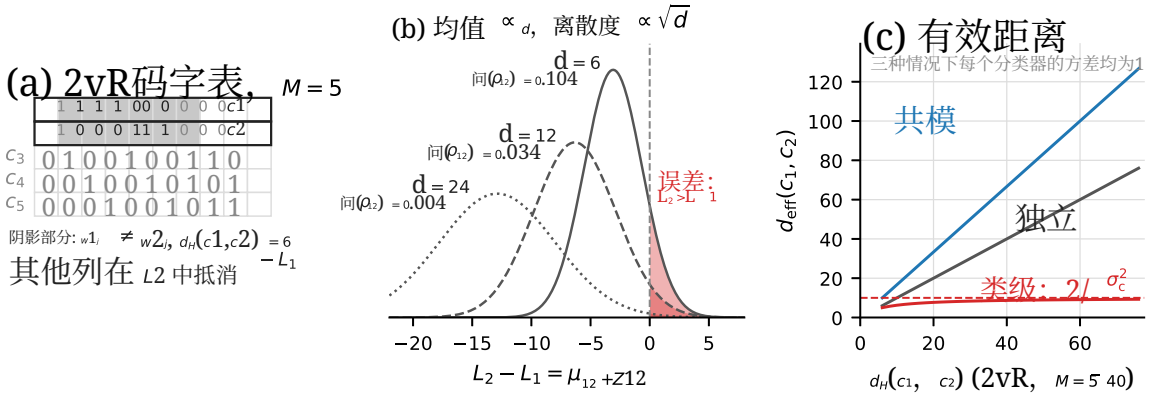


图1：最小距离为何决定超额错误率。(a) 在2vR码字表（$M=5$）中，只有 $d_H(c_1,c_2)=6$ 列（其中 c_1 和 c_2 的码字不同）会进入式（14）的统计量 L_2-L_1；其余列对两个分数的贡献相同。(b) 距离为 $d=6$的2vR编码的成对检验 $L_2-L_1=\mu_{12}+Z_{12}$：$12,24$（$M=5,8,14$；$\theta_1=0.5,\theta_2=0.25$，其余均匀，$\sigma=1$）。均值随 d线性增长，标准差则按$\sqrt{d}$增长，因此错误概率 $Q(\rho_{12})$（红色）随d呈指数衰减。(c) 第8节中沿2vR族的有效距离，比较三种噪声模型；每个分类器的方差均为1：独立噪声（$\sigma^2=1$）、共模噪声（$\sigma_0^2=0.6$，$\sigma_z^2=0.4$）和类别级噪声（$\sigma_0^2=0.6$,$\sigma_c^2=0.2$）。只有类别级噪声会使 d_{eff} 趋于饱和，饱和值为 $2/\sigma_c^2$。

$n_{\min}$ 的定义要求 W 的第 1 行不恒等于 0；对于每个带有 $n<M/2$ 的 nvR 编码，这一条件都成立（对于 nvR 编码，我们使用不涉及 $n_{\min}$ 的推论 1），而对于一般码字表而言，这是一个额外的假设。

命题 3（μ_{1k}的上界：一般码字表）。对于 $\boldsymbol{\theta}\in\Omega(\lambda)$，

$$\mu_{1k}\le d_H(c_1,c_k)\ln\left(\frac{1-\theta_1-\omega(n_{\min})}{\theta_1+\omega(n_{\min})}\right). \tag{24}$$

证明。由（3）可得，$f_j^*(w_{1j}|\boldsymbol{\theta})=\theta_1+\alpha_j$。我们证明对于所有 $j\in\mathcal{F}$，$\alpha_j\ge\omega(n_{\min})$ 均成立。在包含 $w_{1j}=1$ 的列中，α_j 是正侧除 c_1 之外的（n_j-1）个类别的后验概率之和（其中 $n_j\ge n_{\min}$），并且至少等于 $\omega(n_{\min})$，即（$n_{\min}-1$）个最小后验概率之和。在包含 $w_{1j}=0$ 的列中，α_j 是负侧除c_1 之外的（$M-n_j-1$）个类别的后验概率之和；根据码字表的性质（每列中的 1 的数量至多为 $M/2$），其中包含 $M-n_j-1\ge M/2-1\ge n_{\min}-1$ 项，因此同样至少等于 $\omega(n_{\min})$。由于

$$\frac{f_j^*(w_{kj}|\boldsymbol{\theta})}{f_j^*(w_{1j}|\boldsymbol{\theta})}=\frac{1-\theta_1-\alpha_j}{\theta_1+\alpha_j}\le\frac{1-\theta_1-\omega(n_{\min})}{\theta_1+\omega(n_{\min})} \tag{25}$$

对于 $w_{1j}\ne w_{kj}$，(25) 的左侧在 $w_{1j}=w_{kj}$ 时等于 1，将这些代入 $\mu_{1k}=\sum_{j:\,w_{1j}\ne w_{kj}}\ln\frac{1-\theta_1-\alpha_j}{\theta_1+\alpha_j}$，即可得到 (24)。□

对于 nvR 码，利用对结构可以得到一个更紧的界。对于一对满足 $w_{1j}=w_{kj'}\ne w_{kj}=w_{1j'}$ 和对所有 $l\ne 1,k$ 均满足 $w_{lj}=w_{lj'}$ 的分类器 f_j 和 $f_{j'}$，定义 $\eta_j(\alpha_j;\theta_1,\theta_k)\triangleq\frac{f_j^*(w_{kj}|\boldsymbol{\theta})}{f_j^*(w_{1j}|\boldsymbol{\theta})}\cdot\frac{f_j^*(w_{kj'}|\boldsymbol{\theta})}{f_j^*(w_{1j'}|\boldsymbol{\theta})}$。

引理 1（η_j 的上界）。若 $\theta_1>\theta_k$，则在定义域 $\alpha_j\in[0,1-\theta_1-\theta_k]$ 上（由于 $\theta_1+\theta_k\le 1$，该定义域非空）

$$\eta_j(\alpha_j;\theta_1,\theta_k)\le\left(\frac{1-\theta_1+\theta_k}{1+\theta_1-\theta_k}\right)^2, \tag{26}$$

等号在 $\alpha_j = (1-\theta_1-\theta_k)/2$ 处成立。

仅作为 α_j 的函数时， η_j 呈单峰形，并在内部点 $\alpha_j = (1-\theta_1-\theta_k)/2$处达到最大值（附录 A）。

命题4 （ μ_{1k}的上界： nvR 编码）。 当码字表是一个 nvR 编码时，

$$\mu_{1k} \le d(n,M)\ln\!\left(\frac{1-\theta_1+\theta_k}{1+\theta_1-\theta_k}\right). \tag{27}$$

证明。由于一个 nvR 编码包含所有划分为 n 与（$M-n$）的划分，因此，对于任意满足 $w_{1j}\neq w_{kj}$ 的 f_j ，码字表中也存在一个划分，它仅交换 c_1 和 c_k 的位置，同时保持其他类别的归属不变；这给出了引理 1 中指定的对应项 $f_{j'}$。对于 $n = M/2$, 一个将 c_1 与 c_k 分开的列由集合 S 确定，其中 S 是位于 c_1一侧的其他 $n-1$ 个类别的集合；交换 c_1和 c_k后的划分对应于 R \ S，其中 R $=$ C \ {1, k}；这个划分可以作为其补集存储。对列取补集不会改变比值 $f_j^*(w_{kj}|\boldsymbol{\theta})/f_j^*(w_{1j}|\boldsymbol{\theta})$，而映射 S $\mapsto$ R \ S 对于 $n\ge 2$没有不动点，因此配对论证不受影响；命题 6 的证明也同理。根据等距性，带有 $w_{1j}\neq w_{kj}$的列有 $d(n,M)$ 个，构成 $d(n,M)/2$对这样的配对。根据引理 1，每对的贡献 ln η_j至多为 2 ln $\frac{1-\theta_1+\theta_k}{1+\theta_1-\theta_k}$，因此

$$\mu_{1k} = \sum_{\text{pairs}} \ln\eta_j \le \frac{d(n,M)}{2}\cdot 2\ln\frac{1-\theta_1+\theta_k}{1+\theta_1-\theta_k}, \tag{28}$$

这给出了（27）。 □

4.2 超额错误率上界

命题 n 5 （上界：一般码字表）。 对于 $\boldsymbol{\theta}\in\Omega(\lambda)$，当 $\theta_1+\omega(n_{\min})>1/2$ 时，

$$\Delta(\boldsymbol{\theta}) \le (M-1)\exp\!\left(-d_{\min}(c_1)\cdot\kappa_1(\theta_1,\omega(n_{\min}))\right), \tag{29}$$

其中 $\kappa_1(\theta_1,\omega)\triangleq \frac{1}{2\sigma^2}\ln^2\!\left(\frac{\theta_1+\omega}{1-\theta_1-\omega}\right)$。当以 $d_{\min}\le d_{\min}(c_1)$ 替换 $d_{\min}(c_1)$ 时，该上界仍成立。

证明。Un 并集界： $P(E|\boldsymbol{\theta})\le\sum_{k\neq 1}Q(\rho_{1k})$。由命题 3 和 ρ_1 的定义 (9) $_k$，

$$\rho_{1k} \ge \frac{\sqrt{d_H(c_1,c_k)}}{\sigma}\ln\!\left(\frac{\theta_1+\omega(n_{\min})}{1-\theta_1-\omega(n_{\min})}\right). \tag{30}$$

当 $\theta_1+\omega(n_{\min})>1/2$ 时，我们有 $\rho_{1k}>0$，因此应用 Chernoff界 $Q(x)\le e^{-x^2/2}$ 并使用$d_H(c_1,c_k)\ge d_{\min}(c_1)\ge d_{\min}$，得到 (29)。 □

推论 1 （上界： nvR 编码）。 对于 $\boldsymbol{\theta}\in\Omega(\lambda)$,

$$\Delta(\boldsymbol{\theta}) \le (M-1)\exp\!\left(-d(n,M)\cdot\kappa_2(\lambda,\sigma)\right), \qquad \kappa_2(\lambda,\sigma)\triangleq\frac{1}{2\sigma^2}\ln^2\!\left(\frac{1+\lambda}{1-\lambda}\right). \tag{31}$$

证明。 在命题4中，利用 ln $\frac{1+u}{1-u}$ ($u=\theta_1-\theta_k$) 和 $\theta_1-\theta_k\ge\lambda$的单调递增性，

$$\rho_{1k} \ge \frac{\sqrt{d(n,M)}}{\sigma}\ln\!\left(\frac{1+\lambda}{1-\lambda}\right)\triangleq\rho^{\mathrm{L}}. \tag{32}$$

应用并集界和 Chernoff界可得 (31)。 □

并集界与 Gallager 的方法（Gallager，1968年）相呼应：只要距离项 $\kappa\, d_{\min}(c_1)$ 增长得快于竞争者数量的对数 ln$(M-1)$，超额错误率就会趋于零；该对数起到码率的作用。指数是否精确，需通过下界（第5节）和紧性论证（第6节）来确定。

5 下界：汉明距离的必要性

策略。 在等效信道中，与最难竞争者进行比较所产生的错误为$P(E|\boldsymbol{\theta})$提供下界（引理3）；这对任意码字表都成立，并且结合对 μ_{1k} 的逐列界，即可得出必要性（推论 2）。对于具有 $n \geq 2$ 的 nvR 码，对 μ_{1k} 的下界（命题 6）可给出形式为 $Q(\rho^{\mathrm{U}})$ 的显式下界（命题 7）。

引理 2 （Q 函数的双侧评估）。 对于 $x > 0$,

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{x}{x^2+1} \cdot e^{-x^2/2} \leq Q(x) \leq e^{-x^2/2}, \tag{33}$$

对于 $x \geq 1$,

$$\frac{1}{2\sqrt{2\pi}\, x} \cdot e^{-x^2/2} \leq Q(x) \leq e^{-x^2/2}. \tag{34}$$

证明。 上界由 Chernoff界给出。下界源自关于 Mills比的经典结果（Gordon，1941年；Feller，1968年）：$Q(x) \geq \frac{1}{\sqrt{2\pi}} \cdot \frac{x}{x^2+1} e^{-x^2/2}$ $(x > 0)$，由此得到（33）。对于 $x \geq 1$，我们有 $\frac{x}{x^2+1} \geq \frac{1}{2x}$（$\Leftrightarrow 2x^2 \geq x^2 + 1 \Leftrightarrow x^2 \geq 1$），因此（33）给出（34）。 □

5.1 任意码字表的下界及必要性

Let

$$\rho_{\min} \triangleq \min_{k \neq 1} \rho_{1k}. \tag{35}$$

引理3 （ $P(E|\boldsymbol{\theta})$ 的下界（球堆积类比））。 在假设1下，对于任意码字表,

$$P(E|\boldsymbol{\theta}) \geq Q(\rho_{\min}), \tag{36}$$

并且，当 $\rho_{\min} \geq 1$,

$$P(E|\boldsymbol{\theta}) \geq \frac{1}{2\sqrt{2\pi}\, \rho_{\min}} \exp\left(-\frac{\rho_{\min}^2}{2}\right). \tag{37}$$

证明。 根据事件 E 的定义 (15)，$\{R_{1k} \leq 1\} \subseteq E$ 对任意 $k \neq 1$ 都成立，因此

$$P(E|\boldsymbol{\theta}) \geq \max_{k \neq 1} P(R_{1k} \leq 1|\boldsymbol{\theta}) = \max_{k \neq 1} Q(\rho_{1k}) = Q(\rho_{\min}), \tag{38}$$

最后一个等式由 Q 单调递减得出。当 $\rho_{\min} \geq 1$ 时，应用引理 2 的下界 (34) 可得 (37)。

□

该界的作用类似于球堆积界，因为无法避免相对于最近竞争者的误差；但与信道编码中的球堆积界不同，它是一个成对界，且不涉及对译码区域的计数。

注 2 （$\rho_{\min}$ 对于等距码）。 对于等距码， $d_H(c_1, c_k)$ 不依赖于 k，因此$\rho_{1k} \propto |\mu_{1k}|$ 且 $\rho_{\min}$ 由具有最小 $|\mu_{1k}|$ 的竞争类别达到。在有限 M 下，这未必是 c_2，但 $\rho_{\min} \leq \rho_{12}$ 始终成立，因此下界以下均以 c_2 表述；对于 nvR 码，在第 6 节（附录 F）的渐近区域中，对所有充分大的 M， $\rho_{\min} = \rho_{12}$。

推论 2 (与最可能竞争者之间距离的必要性**).** 对于 $k \neq 1$ let $B_k(\boldsymbol{\theta}) \triangleq \ln \frac{1-\theta_k}{\theta_k}$，并令 $B(\boldsymbol{\theta}) \triangleq B_M(\boldsymbol{\theta}) = \max_{k \neq 1} B_k(\boldsymbol{\theta})$。对于任意码字表和任意 $\boldsymbol{\theta} \in \Omega(\lambda)$,

$$\Delta(\boldsymbol{\theta}) \geq \lambda\, Q(\rho_{\min}) \geq \lambda\, Q\left(\min_{k \neq 1} \frac{\sqrt{d_H(c_1, c_k)}\, B_k(\boldsymbol{\theta})}{\sigma}\right) \geq \lambda\, Q\left(\frac{\sqrt{d_{\min}(c_1)}\, B(\boldsymbol{\theta})}{\sigma}\right) > 0. \tag{39}$$

特别地， $\Delta(\boldsymbol{\theta}) \geq \lambda\, Q\big(\sqrt{d_H(c_1,\ c_2)}\ B_2(\boldsymbol{\theta})/\sigma\big)$，其仅通过 λ 和 θ_2依赖于 $\boldsymbol{\theta}$：只要两个最可能类别的码字之间的距离有界，超额错误率就至少为一个正的常数，该常数既不依赖于码长，也不依赖于 M。

证明。 第一个不等式由 $\Delta(\boldsymbol{\theta}) \geq \lambda P(E|\boldsymbol{\theta})$（命题 2）和引理3得到。对于第二个不等式，固定$k \neq 1$ 以及满足 $w_{1j} \neq w_{kj}$的列 j。包含 c_k 的一侧的后验质量为 $f_j^*(w_{kj}|\boldsymbol{\theta}) \in [\theta_k, 1-\theta_1]$，而包含 c_1的一侧的质量为 $f_j^*(w_{1j}|\boldsymbol{\theta}) = 1 - f_j^*(w_{kj}|\boldsymbol{\theta}) \in [\theta_1, 1-\theta_k]$，因此

$$\frac{\theta_k}{1-\theta_k} \leq \frac{f_j^*(w_{kj}|\boldsymbol{\theta})}{f_j^*(w_{1j}|\boldsymbol{\theta})} \leq \frac{1-\theta_1}{\theta_1}. \tag{40}$$

由于 $\theta_k < \theta_1$ 和 $\theta_1 + \theta_k \leq 1$， $\left|\ln \frac{1-\theta_1}{\theta_1}\right|$ 和 $\ln \frac{1-\theta_k}{\theta_k}$均不超过 $B_k(\boldsymbol{\theta})$，且每个此类列的对数比都属于 $[-B_k(\boldsymbol{\theta}), B_k(\boldsymbol{\theta})]$。 $w_{1j} = w_{kj}$ 的列不会对 μ_{1k}产生贡献，因此

$$|\mu_{1k}| \leq d_H(c_1, c_k)\, B_k(\boldsymbol{\theta}) \tag{41}$$

and

$$\rho_{\min} \leq \rho_{1k} \leq \frac{|\mu_{1k}|}{\sigma\sqrt{d_H(c_1,c_k)}} \leq \frac{\sqrt{d_H(c_1,c_k)}\, B_k(\boldsymbol{\theta})}{\sigma} \tag{42}$$

对每个 $k \neq 1$均成立；（39）中的第二个不等式由 Q单调递减得出。第三个不等式通过取 $k = \arg\min_{k' \neq 1} d_H(c_1, c_{k'})$并使用 $B_k(\boldsymbol{\theta}) \leq B_M(\boldsymbol{\theta})$得出，因为 $\theta_k \geq \theta_M$且 $\ln \frac{1-t}{t}$关于 t递减。 □

5.2 nvR 码的显式下界 [$n \geq 2$]

对于满足 $n \geq 2$ 的 nvR 码，我们将 $\rho_{\min}$ 作为 $\boldsymbol{\theta}$的显式函数来求值，这需要对 μ_{1k} 给出下界（即对 $|\mu_{1k}|$给出上界）。

Le引理 4 （ η的下界）。 对于与引理 1 中相同的对（f_j, $f_{j'}$），如果 $\theta_1 > \theta_k$ 该 n

$$\eta_j(\alpha_j;\ \theta_1,\ \theta_k) \geq \frac{(1-\theta_1)\ \theta_k}{\theta_1\ (1-\theta_k)}, \tag{43}$$

在两个端点处均取等号 $\alpha_j = 0$ 和 $\alpha_j = 1 - \theta_1 - \theta_k$。

证明概要。 当 $\theta_1 > \theta_k$， η_j关于 $[0, 1-\theta_1-\theta_k]$ 在 α_j上是单峰的（上升至引理 1 的内部最大值，之后下降）时，因此其最小值在端点处取得，且两个端点处的函数值均等于 $\frac{(1-\theta_1\ \theta_k\ )}{\theta_1\ 1-\theta_k}$（）（附录 B）。 □

命题 6 （ μ_{1k}vR 编码的下界， $n \geq 2$）。 当码字表是一个 nvR 编码时，

$$\mu_{1k} \geq \frac{d(n,M)}{2} \ln\left(\frac{(1-\theta_1)\theta_k}{\theta_1(1-\theta_k)}\right). \tag{44}$$

证明。 根据 nvR 编码的等距性和对结构，

$$\frac{g^*(c_k|\boldsymbol{\theta})}{g^*(c_1|\boldsymbol{\theta})} = \prod_{\text{pairs}} \eta_j(\alpha_j; \theta_1, \theta_k). \tag{45}$$

配对数为 $d(n,\ M)/2$，因此应用引理 4，

$$\frac{g^*(c_k|\boldsymbol{\theta})}{g^*(c_1|\boldsymbol{\theta})} \geq \left(\frac{(1-\theta_1)\theta_k}{\theta_1(1-\theta_k)}\right)^{d(n,M)/2}. \tag{46}$$

取对数可得（44）。 □

With

$$\rho_{1k}^{\mathrm{U}} \triangleq \frac{\sqrt{d(n,M)}}{2\sigma}\left|\ln\frac{(1-\theta_1)\theta_k}{\theta_1(1-\theta_k)}\right|, \tag{47}$$

命题 6 给出 $\rho_{1k} \leq \rho_{1k}^{\mathrm{U}}$。

命题 7（超额错误率下界： nvR 编码， $n \geq 2$）。 当码字表是一个带有 $n \geq 2$ 的 nvR 编码时，对于 $\boldsymbol{\theta} \in \Omega(\lambda)$,

$$\Delta(\boldsymbol{\theta}) \geq \lambda\, Q(\rho_{12}^{\mathrm{U}}), \tag{48}$$

并且，当 $\rho_{12}^{\mathrm{U}} \geq 1$ 时,

$$\Delta(\boldsymbol{\theta}) \geq \frac{\lambda}{2\sqrt{2\pi}\,\rho_{12}^{\mathrm{U}}} \exp\!\left(-\frac{(\rho_{12}^{\mathrm{U}})^2}{2}\right), \qquad \frac{(\rho_{12}^{\mathrm{U}})^2}{2} = d(n,M)\cdot\frac{1}{8\sigma^2}\ln^2\!\left(\frac{\theta_1(1-\theta_2)}{(1-\theta_1)\theta_2}\right). \tag{49}$$

证明。 根据命题 2， $\Delta(\boldsymbol{\theta}) \geq \lambda P(E|\boldsymbol{\theta})$，以及 $\{R_{12} \leq 1\} \subseteq E$给出 $P(E|\boldsymbol{\theta}) \geq Q(\rho_{12})$。由 $\rho_{12} \leq \rho_{12}^{\mathrm{U}}$以及 Q的单调递减性，我们得到（48）。公式（49）由引理 2 的下界（34）推出。（49）中（ρ_{12}^{U}）$^2/2$ 的表达式可通过代入（47）得到。 □

6 紧致性：可靠性函数

本节中的所有结果均限于具有 $n \geq 2$ 的 nvR 码。

策略。 对于等距码，引理3和并集界给出 $Q(\rho_{12}) \leq P(E|\boldsymbol{\theta}) \leq (M-1)Q(\rho_{\min})$，并且对于所有充分大的 M，$\rho_{\min} = \rho_{12}$ 成立，因此除去前因子后，上界和下界指数具有相同的值 $\rho_{12}^2/2$。紧性并非来自令 μ_{1k}上的界相等（它们的系数 $d(n,\ M)$ 和 $d(n,\ M)/2$ 不同）；而是源于 μ_{12}的精确渐近值（引理 6）， $\mu_{12}/(d(n,\ M)/2)\ln\frac{(1-\theta_1\ \theta_2\ )()}{\theta_1\ 1-\theta_2}$，因此两个指数都收敛于

$$e_{\mathrm{lower}} \triangleq \frac{1}{8\sigma^2}\ln^2\frac{\theta_1(1-\theta_2)}{(1-\theta_1)\theta_2} \tag{50}$$

（命题 8）；下标反映了 e_{lower} 首次作为下界 (49) 的指数出现。**引理 5**（ n关于 $P(E|\boldsymbol{\theta})$ 的夹逼： nvR 编码， $n \geq 2$）。 当码字表是一个 nvR 编码时，对于 $\boldsymbol{\theta} \in \Omega(\lambda)$,

$$Q\!\left(\frac{\sqrt{d(n,M)}}{2\sigma}\left|\ln\frac{(1-\theta_1)\theta_2}{\theta_1(1-\theta_2)}\right|\right) \leq P(E|\boldsymbol{\theta}) \leq (M-1)\exp\!\left(-\frac{d(n,M)}{2\sigma^2}\ln^2\!\left(\frac{1+\lambda}{1-\lambda}\right)\right). \tag{51}$$

证明。 下界：由命题 6， $\rho_{1k} \leq \rho_{1k}^{\mathrm{U}}$ （式 (47)）。根据 $P(E|\boldsymbol{\theta}) \geq Q(\rho_{12}) \geq Q(\rho_{12}^{\mathrm{U}})$（$Q$ 单调递减），可得 (51) 的左侧不等式。

上界：由 (32)、并集界和 Chernoff界可得 (51) 的右侧不等式，证明与推论 1 中相同。

□

注 3（为何此处最小距离已足够）。 (51) 中的因子 $M-1$ 来自对 $M-1$ 个竞争码字取并集界，每个竞争码字的距离均为 d ——对于等距码而言这是精确的。对于加性高斯白噪声信道上的等距信号集，例如正交信号和单纯形信号，经典结论表明，该并集界与成对下界具有相同的指数（Viterbi 和 Omura，1979年；Gallager，1968年）；但与那些情形不同的是，此处的信号 μ_{1k} 由后验分布而非固定信号能量决定，并且因竞争码字而异，而指数则由 μ_{12} 的渐近值给出（引理 6）。对于一般码，并集界为 $\sum_\delta A_\delta\, Q(\rho(\delta))$，其中 A_δ 是与 c_1 的距离为 δ 的竞争者数量，而 $(M-1)Q(\rho_{\min})$ 在只有一个竞争码字接近时，可能会松弛至多 M 倍。因此， nvR 限制对应于距离谱退化的情形，此时最小距离足以完整概括。

引理 6（ μ_{12}的渐近精确值： nvR 编码）。 对于满足 $n \geq 2$ 的 nvR 编码，固定 n ，并令 $M \to \infty$。假设 θ_1, θ_2 独立于 M 固定（其中 $1-\theta_1-\theta_2 > 0$），且其余类别的后验概率满足 $\theta_i = O$（$1/M$）（$i \geq 3$）；均匀分配 $\theta_i =$（$1-\theta_1-\theta_2$）/（$M-2$）就是一个例子。于是

$$\frac{\mu_{12}}{d(n,M)/2} \to \ln\frac{(1-\theta_1)\theta_2}{\theta_1(1-\theta_2)} \quad (M \to \infty), \tag{52}$$

更确切地说，

$$\left|\frac{\mu_{12}}{d(n,M)/2}-\ln\frac{(1-\theta_1)\theta_2}{\theta_1(1-\theta_2)}\right|\le\frac{\beta(\theta_1,\theta_2,n)}{M-2} \tag{53}$$

对于每个 $M\ge 2n$，其中 $\beta(\theta_1,\ \theta_2,\ n)\triangleq\left(\frac{1}{\theta_1}+\frac{1}{\theta_2}-\frac{1}{1-\theta_1}-\frac{1}{1-\theta_2}\right)(n-1)(1-\theta_1-\theta_2)$。界（53）仅使用了 $\sum_{i\ge 3}\theta_i=1-\theta_1-\theta_2$，而非 $\theta_i=O(1/M)$；此外，若 $\theta_i>0$对所有 $i\ge 3$均成立，则 $\mu_{12}/(d(n,\ M)/2)$ 对每个 $M\ge 2n$ 都严格大于其极限。

证明略。每个分类器对的对数比 $\gamma(\alpha_{\mathcal{S}})=\ln\frac{(1-\theta_1-\alpha_{\mathcal{S}})(\theta_2+\alpha_{\mathcal{S}})}{(\theta_1+\alpha_{\mathcal{S}})(1-\theta_2-\alpha_{\mathcal{S}})}$满足 $\gamma(0)\le\gamma(\alpha_{\mathcal{S}})\le\gamma(0)+c'\alpha_{\mathcal{S}}$，其中 $c'=1/\theta_1+1/\theta_2-1/(1-\theta_1)-1/(1-\theta_2)$；对跨越这些分类器对的子集 $\mathcal{S}$求平均，其均值 $\bar{\alpha}_{\mathcal{S}}$ 为 $(n-1)(1-\theta_1-\theta_2)/(M-2)$，即可得出该论断（附录 E ）。 □

定义 6(ECOC 可靠性函数)。对于一个 nvR 编码，其中 n 固定且 $M\to\infty$（因此 $d(n,\ M)\to\infty$），

$$E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sigma)\triangleq-\lim_{d(n,M)\to\infty}\frac{\ln P(E|\boldsymbol{\theta})}{d(n,M)}. \tag{54}$$

由于 $\boldsymbol{\theta}$ 的维数为 M，因此将其理解为一个序列 $\boldsymbol{\theta}^{(M)}\in\Omega(\lambda)$，其中 $\theta_1^{(M)}=\theta_1$ 和 $\theta_2^{(M)}=\theta_2$ 固定，且 $\theta_i^{(M)}=O(1/M)$ $(i\ge 3)$，即引理 6 中的族；命题 8 表明，极限与具体序列无关。

命题 8（可靠性函数的存在性和值）。对于一个 nvR 编码，若 $n\ge 2$ 且取上述族中的任意序列 $\boldsymbol{\theta}^{(M)}$，则 ECOC 可靠性函数存在，且

$$E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sigma)=e_{\mathrm{lower}}=\frac{1}{8\sigma^2}\ln^2\left(\frac{\theta_1(1-\theta_2)}{(1-\theta_1)\theta_2}\right). \tag{55}$$

证明概要。由引理 6，$\rho_{12}^2/(2d(n,\ M))\to e_{\mathrm{lower}}$，且对所有充分大的 M，$\rho_{\min}=\rho_{12}$。由并集界和 Chernoff 界可得 $\liminf\ {}^{-\ln P(E|\boldsymbol{\theta})}/_{d}\ge e_{\mathrm{lower}}$；由引理 3 和引理 2 可得 $\limsup\ {}^{-\ln P(E|\boldsymbol{\theta})}/_{d}\le e_{\mathrm{lower}}$，因此极限存在且等于 e_{lower}（附录 F）。 □

Co推论 3（超额错误率的可靠性函数）。对于一个 nvR 编码，满足 $n\ge 2$,

$$-\lim_{d(n,M)\to\infty}\frac{\ln\Delta(\boldsymbol{\theta})}{d(n,M)}=E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sigma). \tag{56}$$

证明。对命题 2 的双侧评估 $\lambda P(E|\boldsymbol{\theta})\le\Delta(\boldsymbol{\theta})\le\theta_1 P(E|\boldsymbol{\theta})$ 取对数，并除以 $d(n,\ M)$,

$$\frac{-\ln(\theta_1 P(E|\boldsymbol{\theta}))}{d(n,M)}\le\frac{-\ln\Delta(\boldsymbol{\theta})}{d(n,M)}\le\frac{-\ln(\lambda P(E|\boldsymbol{\theta}))}{d(n,M)}. \tag{57}$$

由于 $\lambda,\ \theta_1$ 是与 $d(n,\ M)$ 无关的正常数，因此当 $d(n,\ M)\to\infty$时，我们有 $-\ln\lambda/d(n,\ M)\to 0$，并且 $-\ln\theta_1/d(n,\ M)\to 0$。于是，当 $-\ln P(E|\boldsymbol{\theta})/d(n,\ M)$ 时，两侧具有相同的极限，由命题 8 可得（56）。 □

推论 1 中的系数 $\kappa_2(\lambda,\ \sigma)$ 在 $\Omega(\lambda)$ 上均匀成立，而 $E_{\mathrm{ECOC}}(\boldsymbol{\theta},\ \sigma)$ 是每个 $\boldsymbol{\theta}$ 处的精确指数；二者关系如下。

引理 7（一致指数与逐点指数）。对于任意 $\boldsymbol{\theta}\in\Omega(\lambda)$，有 $E_{\mathrm{ECOC}}(\boldsymbol{\theta},\ \sigma)\ge\kappa_2(\lambda,\ \sigma)$，当 $\theta_1+\theta_2\to 1$，$\theta_1-\theta_2=\lambda$ 时取极限等号；即 $\kappa_2(\lambda,\ \sigma)=\inf_{\boldsymbol{\theta}\in\Omega(\lambda)}E_{\mathrm{ECOC}}(\boldsymbol{\theta},\ \sigma)$（附录 C）。

引理 6 的收敛性（ $O(1/M)$ 码率，$M=4$–500）、蒙特卡洛指数$-\ln P(E|\boldsymbol{\theta})/d(n,\ M)$ 收敛至 E_{ECOC}，以及引理 5 与蒙特卡洛估计（$4\le M\le 20$）的一致性，均已在附录 H 中通过数值验证。在那里考察的所有情形（$n=2$）中，有限长度指数都超过了 e_{lower}。这种超出源于 $P(E|\boldsymbol{\theta})$ 的次指数前因子；引理 6 的校正项则起相反作用，因为 $r(M)>0$意味着 $\rho_{12}^2/(2d(n,\ M))<e_{\mathrm{lower}}$。因此，这种超出并非

一般情况下都能保证：对于 $n \geq 3$ 以及定义 6 的族， $d(n,\ M) = \Theta(M^{n-1})$ 增长得足够快，以至于 $O(1/M)$ 修正项超过前因子的影响；对于足够大的 M，引理 3 结合（33）给出 $-\ln P(E|\boldsymbol{\theta})/d(n,\ M) < e_{\text{lower}}$（在附录 H 的设定中，适用于 $n = 3$ 和 $M \geq 25$）。

7 主要定理和特定编码

定理1（根据最小距离刻画超额错误率）。 在假设1下，对于$\boldsymbol{\theta} \in \Omega(\lambda)$（$\lambda > 0$）：

1. 充分性（命题 5、推论 1）：对于满足命题 5 条件的任何码字表以及 $\kappa = \kappa_2(\lambda, \sigma)$的 nvR 码，$\Delta(\boldsymbol{\theta}) \leq (M - 1)\exp(-\kappa\, d_{\min}(c_1))$，其中 $\kappa = \kappa_1(\theta_1, \omega(n_{\min}))$。由于二元分类器彼此不同，因此 $N < 2^M$，所以在固定的 M 下，$d_{\min}(c_1) \leq N$ 有界，并且只能与 M一起增长；沿着码字表的任意序列，当 $\kappa\, d_{\min}(c_1) - \ln(M-1) \to \infty$时，$\Delta(\boldsymbol{\theta}) \to 0$。对于满足 $n \geq 2$ 的 nvR 码，由于 $M \to \infty$，这对Ω(λ)一致成立，因为ln$(M-1) = o(d(n, M))$。
2. **必要性**（推论 2）：对于任意码字表，$\Delta(\boldsymbol{\theta}) \geq \lambda\, Q\big(\sqrt{d_H(c_1,\ c_2)}\ B_2(\boldsymbol{\theta})/\sigma\big) > 0$。因此，如果 $d_H(c_1,\ c_2)$（特别是对于等距码，如果 $d_{\min}(c_1)$）保持有界，则超额错误率将被一个由 λ、θ_2、σ 及该距离决定的正常数从下方界定，并且这一界定对 N 和 M 均一致成立。

3. **紧性**（命题 8）：对于具有 $n \geq 2$ 的 nvR 码，上、下指数重合，且可靠性函数为 $E_{\text{ECOC}}(\boldsymbol{\theta},\ \sigma) = \frac{1}{8\sigma^2}\ln^2\!\left(\frac{\theta_1}{(1-\theta_2)}\ \frac{(1-\theta_1)}{\theta_2}\right)$。

定理1给出了固定编码下信道编码定理各部分的ECOC对应形式：充分性对应可达性，必要性对应逆定理，而紧性对应可靠性函数。对于随机抽取的码，相应的表述是Kobayashi 等人（2024）的平均值界，其中码长通过码率 $\ln M/N$ 体现；在定理1中，N的作用由 $d_{\min}(c_1)$承担。这正是ECOC码所处的情形：对于具有 $n \geq 2$ 的 nvR码，以及对于哈达玛码，码率 $\ln M/N$ 趋于零；在低码率下，信道编码的错误指数已经由在最小距离处混淆两个码字的概率决定，而不是由码长决定（Gallager，1965年，第五节）。必要性对每个码字表都成立，充分性对满足命题 5条件的每个码字表都成立，而对于具有 $n \geq 2$ (n 固定，$M \to \infty$)的 n vR码，可靠性函数的值也得到了证明。假设1在定理1中仅通过（14）式中每个 Z_{1k} 的边缘分布体现；第8节中的定理 2消除了独立性要求。

7.1 特定编码

表 1 总结了固定 θ_1 和 θ_2 时指数 $-\ln \Delta(\boldsymbol{\theta})$ 的阶。对于每一行，由推论 2 和 $k = 2$ 可得一个上阶。对于 2vR、Hadamard 和穷举码，由于其充分性界趋于零，匹配的下阶可由第 4 节的上界推出。对于 1vR 和 logM，最小距离是有界的；由于 $\Delta(\boldsymbol{\theta}) \leq \theta_1$，当推论 2 给出 $O(1)$ 时，该阶为 $\Theta(1)$；而对于 $d_H(c_1, c_2)$ 增长的 logM，目前只有上阶 $O(\log M)$ 可得。

1vR 和 logM。 对于 1vR，$d(1, M) = 2$ and $\mu_{1k} = \frac{\ln(1-\theta_1\ \theta_k)()}{\theta_2 - 1 - \theta_2}$ 恰好为 ($\alpha_j = 0$，引理 4 的等号情形)，因此推论 2 表明，超额错误率关于 M 一致地远离零，可靠性函数没有定义。对于 logM，$d_H(c_1, c_2) \leq N = \lceil \log_2 M \rceil$，所以推论 2 给出 $-\ln \Delta(\boldsymbol{\theta}) = O(\log M)$；当 c_1 和 c_2 的码字之间的距离有界时，超额错误率保持远离零（例如，二进制表示中的相邻索引）。

2vR。由于 $d(2,\ M) = 2(M-2) \to \infty$（对于 $M \geq 5$；$d(2,\ 4)$ 根据 (2) $= 2$），超额错误率以指数速度随 $d(2,\ M)$ 衰减，在 Ω(λ) 上以码率 $\kappa_2(\lambda,\ \sigma)$ 均匀成立（推论 1）；并且在每个 $\boldsymbol{\theta}$ 处，都以精确码率 $E_{\text{ECOC}}(\boldsymbol{\theta},\ \sigma) \geq \kappa_2$（命题 8）。

哈达玛码。哈达玛码（定义 3）在码长为 $N = M - 1$ 时是等距的，且 $d = M/2$；但对于 $M \geq 8$，它不具备引理 1 所定义的对结构，因此命题 4 和命题 6 不适用。

事实上，在 ± 1 的形式下，两列配对列恰好在第 1 行和第 k 行不同，因此它们的逐项乘积恰有 2 个元素等于 -1（若其中一列取补，则为 $M-2$；这不会改变每个比值 $f_j^*(w_{kj}|\boldsymbol{\theta})/f_j^*(w_{1j}|\boldsymbol{\theta})$），而 Hadamard矩阵中不同的列彼此正交，因此乘积恰有 $M/2$ 个这样的元素。仅当 $M=4$ 时，这两个计数才相同；此时哈达玛码在列取补意义下等价于 2vR。对于 $M \geq 8$，没有任何类别对 $(c_1,\ c_k)$ 包含哪怕一列配对列。通过另一种论证，形如 (27) 和 (44) 的界仍可能适用于哈达玛码；这仍是一个开放问题。表中每一列都有 $M/2$ 个 1；在必要时对列取补（这不会改变解码问题）后，第 1 行并非全为 0，因此 $n_{\min}=M/2$。于是，命题 5 给出：只要 $\theta_1+\omega(M/2)>1/2$，就有 $\Delta(\boldsymbol{\theta}) \leq (M-1)\exp\left(-\frac{M}{2}\kappa_1(\theta_1,\ \omega(M/2))\right)$。结合针对 $k=2$ 的推论 2，当 $\theta_1+\omega(M/2) \geq 1/2+c$（其中 $c>0$ 为常数）时，指数在码长为 $O(M)$ 时的阶为 M。例如，如果后验质量在 c_1 和 c_2 之外的部分均匀分布，那么 $\omega(M/2)=(1-\theta_1-\theta_2)/2$，并且对于每个 $\boldsymbol{\theta} \in \Omega(\lambda)$，该条件在 $c=\lambda/2$ 下成立；若这部分质量集中在少数几个类别上，则条件可能不成立，此时条件化简为 $\theta_1 \geq 1/2+c$。它是否能达到可靠性函数，以及其系数是多少，仍是开放问题。

穷举码。 作为所有 nvR 编码的串接，它是等距的，且具有 $d_{\text{exh}}=\sum_{n=1}^{\lfloor M/2 \rfloor} d(n,\ M)=2^{M-2}$。逐块应用 μ_{1k} 界，超额错误率随 M 以双指数方式趋于零，其指数的阶为 $N_{\text{exh}}=2^{M-1}-1$。只有该阶是紧的（系数为 κ_2，而逐点的 $e_{\text{exh}}(\boldsymbol{\theta}) \geq \kappa_2$ 则不同，因为 $\bar{\alpha}_{\mathcal{S}} \to 0$ 在具有 $n \sim M/2$ 的块上失效）；参见推论 4（附录D）。

Table 1：码长、最小距离以及指数 $-\ln\ \Delta(\boldsymbol{\theta})$ 的阶（$\theta_1,\ \theta_2$ 固定 ）.

Code	N	$d_{\min}$	$-\ln\ \Delta(\boldsymbol{\theta})$ 的阶
logM	$\lceil \log_2 M \rceil$	1	$O(\log\ M)$；如果 $d_H(c_1,\ c_2)$ 有界，则为 $\Theta(1)$
1vR	M	2	$\Theta(1)$（无衰减）
2vR	$M(M-1)/2$	$2(M-2)$	$\Theta(M)$；$-\ln\ \Delta(\boldsymbol{\theta}) \sim 2M\,E_{\text{ECOC}}$
Hadamard	$M-1$	$M/2$	$\Theta(M)$，若 $\theta_1+\omega(M/2) \geq 1/2+c$；系数开区间
穷举码	$2^{M-1}-1$	2^{M-2}	$\Theta(2^M)$；系数开放

$O(\log M)$ 对于 logM 仅是上阶项；$c>0$ 是一个与 M 无关的常数；2vR 行假设 $M \geq 5$，其系数对应定义 6 的后验族。

其中，哈达玛码是唯一一种在码长为 $O(M)$ 时达到 M 阶指数的编码（比率为 $d_{\min}/N \approx 1/2$）。

8 相关噪声与有效距离

假设1要求噪声相互独立，但第 3–6 节仅通过式 (14) 中每个 Z_{1k} 的边缘分布使用这一点；该边缘分布是方差为 $d_H(c_1,\ c_k)$ 的中心高斯分布；σ^2的联合分布从未被使用（参见备注 1）。令 $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_N)^\top \sim \mathcal{N}(\mathbf{0},\ \Sigma)$，其协方差为任意的 $\Sigma \succ 0$，并令 $\boldsymbol{s}_k \triangleq (w_{kj}-w_{1j})_{j=1}^N$，从而$Z_{1k}=\boldsymbol{s}_k^\top \boldsymbol{\varepsilon} \sim \mathcal{N}(0,\ \boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k)$。我们将该对 $(c_1,\ c_k)$ 的有效噪声水平和有效距离 定义为

$$\sigma_{\text{eff}}^2(1,k) \triangleq \frac{\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k}{d_H(c_1,c_k)}, \qquad d_{\text{eff}}(c_1,c_k) \triangleq \frac{d_H(c_1,c_k)}{\sigma_{\text{eff}}^2(1,k)} = \frac{d_H(c_1,c_k)^2}{\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k}. \tag{58}$$

在假设1下，$\sigma_{\text{eff}}(1,\ k)=\sigma$ 和 $d_{\text{eff}}(c_1,\ c_k)=d_H(c_1,\ c_k)/\sigma^2$。

定理 2（相关高斯噪声）。令 $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \Sigma)$，其中以 $\Sigma \succ 0$ 替代假设1，并令$\boldsymbol{\theta} \in \Omega(\lambda)$ $(\lambda>0)$。

1. **采用有效距离的定理1。** 定理1的第1部分和第2部分成立，其中用 $d_{\text{eff}}(c_1,\ c_k)$替代 $d_H(c_1,\ c_k)$：

$$\Delta(\boldsymbol{\theta}) \leq (M-1)\exp\Big(-\bar{\kappa}\min_{k\neq 1} d_{\text{eff}}(c_1,c_k)\Big), \qquad \Delta(\boldsymbol{\theta}) \geq \lambda\, Q\Big(\sqrt{d_{\text{eff}}(c_1,c_2)}\,B_2(\boldsymbol{\theta})\Big) > 0, \tag{59}$$

其中 $\bar{\kappa}$ 是定理1在 $\sigma=1$ 处取值的常数 κ，并满足相同条件。对于满足 $n \geq 2$ 的 nvR 编码，以及定义 6 中该族的序列 $\boldsymbol{\theta}^{(M)}$，将 σ^2 替换为 v 后，命题 3 仍成立，

也就是说，当 $\sigma_{\mathrm{eff}}^2(1,\ 2) \to v \in (0,\ \infty)$ 且 $\bar{\max}_{k\neq 1}\sigma_{\mathrm{eff}}^2(1,\ k)$ 随 $M \to \infty$ 保持有界时，$-\ln \Delta(\boldsymbol{\theta})/d(n,\ M) \to E_{\mathrm{ECOC}}(\boldsymbol{\theta},\ \sqrt{v})$。

2. **共模噪声与类别级噪声**。令 $\varepsilon_j = \xi_j + \sum_{i:\, w_{ij}=1}\chi_i + z$ 相互独立，且分别服从 $\xi_j \sim \mathcal{N}(0, \sigma_0^2)$、$\chi_i \sim \mathcal{N}(0, \sigma_{\mathrm{c}}^2)$ 和 $z \sim \mathcal{N}(0, \sigma_{\mathrm{z}}^2)$，其中 $\sigma_0 > 0$。若 $\sigma_{\mathrm{c}} > 0$，则对任意码字表 $d_{\mathrm{eff}}(c_1, c_k)$，对所有 $k \neq 1$ 均有 $\leq 2/\sigma_{\mathrm{c}}^2$，因此无论 N 和 $d_{\min}$ 为何，$\Delta(\boldsymbol{\theta}) \geq \lambda\, Q\big(\sqrt{2}\, B_2(\boldsymbol{\theta})/\sigma_{\mathrm{c}}\big) > 0$。若 $\hat{} W$ 是距离为 d 的等距编码且行权重相等，则 i 的分布与假设1中 $\sigma^2 = \sigma_0^2 + \sigma_{\mathrm{c}}^2 d/2$ 时相同；若 $\sigma_{\mathrm{c}} = 0$，则仅相等行权重就足够。

证明概要。(1) 当 $\rho_{1k} \triangleq -\mu_{1k}/(\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k)^{1/2}$ 时，(11) 仍如前成立；命题 2 以及关于 μ_{1k} 的界与噪声无关，而并集界和引理3仅用到每个 Z_{1k} 的分布。(2) 此处 $\Sigma = \sigma_0^2 I + \sigma_{\mathrm{c}}^2 W^\top W + \sigma_{\mathrm{z}}^2 \mathbf{1}\mathbf{1}^\top \succeq \sigma_{\mathrm{c}}^2 W^\top W$，且 $d_H(c_1, c_k) = (\boldsymbol{e}_k - \boldsymbol{e}_1)^\top W \boldsymbol{s}_k \leq \sqrt{2}\,\|W \boldsymbol{s}_k\|$，因此 $\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k \geq \sigma_{\mathrm{c}}^2 d_H(c_1, c_k)^2/2$。详细内容见附录G。 □

情形 $\sigma_{\mathrm{c}} = 0$ 是等相关噪声，即 $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2[(1-\tau)I + \tau \mathbf{1}\mathbf{1}^\top]$，其中 $\sigma_0^2 = (1-\tau)\sigma^2$ 且 $\sigma_{\mathrm{z}}^2 = \tau\sigma^2$：所有分类器共有的相关性会在每个得分差中抵消，并且只会降低有效噪声水平。具有 $n < M/2$ 的 nvR 码具有相等的行权重 $\binom{M-1}{n-1}$，因此对它们而言，定理1适用于 $(1-\tau)\sigma^2$。类别级噪声的表现则有所不同。对于具有 $n < M/2$ 的 nvR 码，每个分类器都有相同的边际方差 $\sigma_0^2 + n\sigma_{\mathrm{c}}^2 + \sigma_{\mathrm{z}}^2$，且与 M 无关；然而，随着 $d(n, M)$增大，$d_{\mathrm{eff}} = d/(\sigma_0^2 + \sigma_{\mathrm{c}}^2 d/2)$会在 $2/\sigma_{\mathrm{c}}^2$处饱和，超额错误率不再趋于零（图1(c)）。因此，必须超过 $\ln(M-1)$增长的是有效距离；有效距离结合了码字之间的分离程度与分类器误差的相关性，而汉明距离就是独立噪声下的有效距离值。定理 2涉及解码器（4）。在第2部分中，类别分量通过每一列的正侧引入，但该下限并不取决于这种方向约定：如果每个类别分量都带有其所在侧的符号，即 $\varepsilon_j = \xi_j + \sum_i (2w_{ij} - 1)\chi_i + z$，则使用 $(\boldsymbol{e}_k - \boldsymbol{e}_1)^\top (2W - \mathbf{1}\mathbf{1}^\top)\boldsymbol{s}_k = 2d_H(c_1, c_k)$的相同论证可得 $d_{\mathrm{eff}}(c_1, c_k) \leq 1/(2\sigma_{\mathrm{c}}^2)$。

9 实验

9.1 设置

我们检验两个预测：(i) 在满足假设1的设置中，超额错误率以 d 按 e_{lower}的速率呈指数衰减；(ii) 对于训练后的分类器，超额错误率会随着最小距离的增大而降低；当距离超过适中水平后，性能不再有明显改善。真实数据检验是定性的，因为训练后的分类器并不完全满足假设1，而且尚未建立一种无需真实后验概率即可稳定估计其噪声水平的方法（第10节）。

L0（半合成）设置。将满足假设1且已知 σ 的对数几率噪声添加到2vR编码的真实 f_j^*中，其中 $\theta_1 \in \{0.4,\ 0.5,\ 0.6,\ 0.7\}$，$\theta_2 = 0.25$，其余质量均匀分布，并且 $6 \leq M \leq 24$，使得 $d = 2(M-2)$ 的范围为 8 到 44；通过蒙特卡洛估计 $P(E|\boldsymbol{\theta})$和 $\Delta(\boldsymbol{\theta})$。

真实数据。我们使用 CIFAR-10、CIFAR-100 和 ImageNet-1k 的 ResNet-50 特征（在 ImageNet-1k 上预训练），并通过主成分分析将其降至 100 维。在每组 $S = 40$ 随机种子中，我们抽取 M 个类别（分别为 $M = 8$、12 和 16 个），并对每个类别的 400 张图像（ CIFAR ）或 150 张图像（ ImageNet-1k ）进行训练；测试时使用 CIFAR 测试集和 ImageNet-1k 验证集。每个二元分类器都是一个逻辑回归模型（逆正则化强度为 1.0），其输出 p_j 估计 $f_j(1|\boldsymbol{x})$。为在固定 M 的情况下改变 $d_{\min}$，我们从固定的 2vR 列和 3vR 列池中抽取 K 列的子集（对于 $M = 8$ 和 12，分别使用全部 84 列和 286 列；对于 $M = 16$，使用 120 个 2vR 列和随机选取的 200 个 3vR 列），并记录每个子集的最小距离；带有 $d_{\min} = 1$ 的点表示这些子集，而非 logM 码。ECOC预测采用求和解码 $\hat{\imath}_{\mathrm{sum}} = \arg\max_{i \in \mathcal{C}} \sum_j [w_{ij} p_j + (1 - w_{ij})(1 - p_j)]$。

参考错误率。对于真实数据，贝叶斯错误率是未知的。作为参考，我们采用基于对数似然（LL）的解码错误率 $\hat{\imath}_{\mathrm{LL}} = \arg\max_{i \in \mathcal{C}} \sum_j [w_{ij} \ln\, p_j + (1 - w_{ij})\ln(1 - p_j)]$，并对整列进行计算

该池对所有子集均固定，并定义经验超额误差 $\Delta_{\mathrm{real}}(d) \triangleq \hat{} P_{\mathrm{err}}^{\mathrm{sum}}(d) - P_{\mathrm{floor}}$。 P_{floor}

是分类器的错误率，因此至少等于贝叶斯错误率，所以 Δ_{real} 低估了真实的超额错误率；它是给定二元分类器上的参考值，而不是贝叶斯错误率。作为比较，在单独的 CIFAR-100 实验中（$M \in \{4、6、8、10、12、16\}$，25 个随机种子），在相同特征上训练的多类逻辑回归，在 150 个条件中的 136 个条件下错误率高于一对其余分类器的 LL 解码，并在 8 个条件下相等（平均差值为 0。011；附录 I）。在同一分类器集合上，对于 $M = 6$–16，求和解码与 LL 解码的差异在 ± 0.005 以内（显著性水平为 0。05 的 TOST等效性检验），并且对于 $M = 4$ 与其完全相同，因此解码规则并非限制因素（附录 I）。

9.2 结果

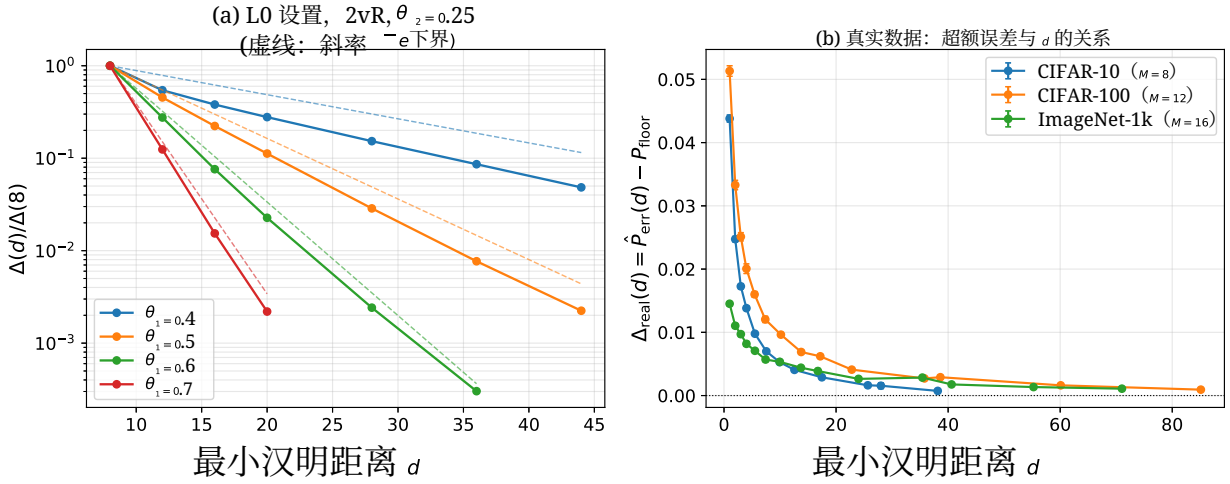


图 2：超额错误率与最小汉明距离 d 的关系。(a) L0 设置（2vR，$\theta_2 = 0.25$）：对数尺度下的 $\Delta(d)/\Delta(8)$；虚线的斜率为 $-e_{\text{lower}}$。(b) 真实数据（ResNet-50 特征，$S = 40$ 个随机种子）：$\Delta_{\text{real}}(d) = \hat{P}_{\text{err}}^{\text{sum}}(d) - P_{\text{floor}}$，按随机列子集的最小距离分箱（误差条：标准误）。这些子集并非等距，因此其最小距离无法完整概括该编码（注 3）。

L0 设置（图 2(a)）。 Δ 随 d呈指数衰减，且快于斜率 $-e_{\text{lower}}$ （拟合斜率为 1.02–1.33 倍的 e_{lower}，最小 θ_1时最大），这与附录 H 的有限长度指数预期一致。对于 2vR 编码，由于 $P(E|\boldsymbol{\theta})$ 的次指数前因子，其有限长度指数从上方趋近于 e_{lower} （附录 H.2）。

真实数据（图 2(b)）。对于全部三个数据集，$\Delta_{\text{real}}(d)$ 会随着 d 增加而下降：对于 CIFAR-10（$M = 8$），从 0.044（$d = 1$）降至 0.0007（$d \approx 38$）；对于 CIFAR-100（$M = 12$），从 0.051 降至 0.0009（$d \approx 85$）；对于 ImageNet-1k（$M = 16$），从 0.015 降至 0.0011（$d \approx 71$）。除了 CIFAR-100 和 ImageNet-1k 各自相邻箱之间约 0.0002 的一次反转外，分箱曲线均单调下降；该反转比箱内标准差小一个数量级（≈ 0.003）。我们将饱和距离 d^*操作性地定义为使 Δ_{real} 的均值落在距零一个箱内标准差以内的最小箱距；据此，对于 CIFAR-10、CIFAR-100 和 ImageNet-1k，该距离分别为 $d^* \approx 38$、36 和 24。这是一个实用阈值：超过 d^* 后，剩余的平均超额误差小于随机种子和列子集造成的离散程度；这并不意味着 Δ_{real} 在统计上与零无法区分（图 2(b) 中的误差条表示标准误）。对于 CIFAR-10（$M = 8$），该标准仅在接近所探测的最大距离处才满足，因此其趋于平稳的现象不如 CIFAR-100 和 ImageNet-1k 明显。超过 d^*后，增加d 对错误率的影响小于随机种子和列子集造成的变化，而分类器数量仍在增长。ImageNet-1k 的初始超额误差要小得多（在 $d = 1$处为 0.015，而参考错误率为 0.017），并且在更小的距离处达到 d^*；ResNet-50 特征是在 ImageNet-1k 标签上训练的，因此能够很好地区分这些类别。对于 $M = 16$，2vR编码具有 $d = 28$，按此标准略高于 d^*，尽管此处的平均超额误差仍约为 0.003。

表2：ImageNet-1k：随着类别数量 M 增加，ˆ P_{err} （ResNet-50 特征、PCA-128， $S = 3$ 个随机种子上的均值 ±标准差）。多类逻辑回归（Softmax）作为参考列出。

M	d_{2vR}	2vR	logM（d=1）	多类逻辑回归（参考 .）
16	28	0.037±0.017	0.126±0.014	0.037±0.017
128	252	0.102±0.003	0.682±0.018	0.121±0.005
256	508	0.128±0.007	0.871±0.009	0.156±0.008

更大的距离并不总是更有帮助。在 CIFAR-100（$M \in \{8{,}10{,}12\}$，25 个随机种子）上，3vR编码的距离 $d(3, M)$ 是 2vR 的 2.5–4.5 倍，但其错误率却略高于 2vR（$+0.0020$，Wilcoxon 检验符号秩检验 $p = 1 \times 10^{-5}$，75 对；附录 I）。第 4–6 节的界限随距离增加而下降，并未解释这一虽小但系统性的性能差距；第10节将通过第8节中的有效距离对此进行考察。

缩放 M。 表2将 ImageNet-1k 的类别数量增加到 $M = 256$ （一次单独的运行：将 PCA 降至 128 维，使用 $S = 3$ 个随机种子，采用 2vR 和 logM 编码本身，而非列子集）。logM 编码（$d = 1$）迅速退化至ˆ $P_{\mathrm{err}} \to 1$，而 2vR（$d = 2(M - 2)$）保持较低水平，正如定理1的必要性和充分性部分所预期的那样；2vR 也与（$M = 16$）多类逻辑回归相当，或低于（$M = 128,\ 256$）多类逻辑回归。本次运行未计算 LL解码参考值，因此该表无法衡量与 P_{floor} 的距离，其 $M = 16$ 值也无法与图 2(b) 进行比较，因为后者使用了不同的训练集大小和列子集。

10 讨论：模型能捕捉什么，不能捕捉什么

该理论能够预测真实数据上的定性行为——随最小距离增加而衰减、短码失效以及趋于平稳——但无法预测其定量形式。四个因素造成了这种差距，并且都能在我们的结果中观察到；第三个因素由定理 2 部分涵盖。

（*i*）后验分布的离散程度。定理1对每个 $\boldsymbol{\theta}$都成立，而ˆ P_{err} 则对输入取平均。在固定的 $\boldsymbol{\theta}$ 下，成对项 $Q(\rho_{1k})$ 除了 $e^{-dE_{\mathrm{ECOC}}(\boldsymbol{\theta},\sigma)}$ 之外，还带有一个阶为 $d^{-1/2}$ 的前因子（引理 2）。对指数 $E_{\mathrm{ECOC}}(\boldsymbol{\theta}, \sigma)$趋近于其下确界的输入取平均，还会引入一个额外的代数因子，因此整体曲线并非单一指数函数。其衰减由指数最小的输入决定，也就是那些接近决策边界的输入。这与间隔条件在插件分类器超额风险中的作用相呼应（Audibert与Tsybakov, 2007）。在我们的曲线上，单一指数函数无法很好地拟合 d （在图 2(b)的各分箱上，对$\ln \Delta_{\mathrm{real}}$ 与 d 进行线性拟合，CIFAR-10、CIFAR-100和ImageNet-1k对应的结果分别为 $R^2 = 0.90$、0.83和0.89），而且无法从这些曲线中读出码率。

(ii) 距离谱。 图 2(b) 的列子集并非等距。对于此类码，错误率取决于整个距离谱 $\{A_\delta\}$ （注 3），而最小距离并不能完整概括其特性。

(iii) 相关噪声与有效距离。根据定理 2，决定相关噪声下超额错误率的量并非 d_H，而是有效距离 $d_{\mathrm{eff}}(c_1,\ c_k) = d_H(c_1,\ c_k)^2/(\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k)$，它通过不同位置的集合取决于编码。定理 2 的类级模型解释了趋于平稳现象，因为随着 d_H 增大， d_{eff} 会趋于饱和；但该模型本身无法解释性能差距，要解释这一点，较长码在关键类别对上的有效距离必须更小。为检验这一点是否能解释第 9.2 节中 3vR 的性能差距，我们使用从校准教师模型重采样的标签，重复了 CIFAR-100 对比实验。这样一来，真实后验概率，以及 Abe & Kumoi (2026) 的 logit 噪声模型下每个二元分类器的对数优势残差，都是已知的；类别、图像和分类器与第 9.1 节相同（附录 I）。在这一设置下，性能差距依然存在（$+0.0012$ LL 解码下的超额错误率，在 75 个种子匹配对中有 57 对为正，Wilcoxon 检验 $p = 5 \times 10^{-7}$）。根据留出残差估计的残差协方差 ˆΣ，在很大程度上可由类别级噪声解释：定理 2 的类别层级模型对其拟合得到 $R^2 = 0.64$–0。83，且其饱和水平 $2/\sigma_c^2$ 接近测得的 d_{eff}，因此两个码都已处于饱和区间。相应地，对于 $M = 8,\ 10,\ 12,$ 3vR 跨类别对的平均有效距离至多是 2vR 的 1。04 倍，其类别对间的最小值仅为 2vR 的 0。89–0。94 倍（跨种子中位数），尽管其汉明距离大出 2。5–4。5 ˆ 倍。带有 Σ 的高斯模型预测 3vR 在每个随机种子上的错误率都更高，而使用拟合的 σ 的假设1则预测其错误率更低；对于置信度较低的教师模型，有效距离之比

上升至 1。22–1。25 时，3vR 的表现略好（-0.0004），这与高斯模型的预测一致（-0.0003）。然而，第一种设置下预测的性能差距（$+0.0074$）约为观测值的六倍；与高斯模型不同，改用重采样的经验残差后，预测的性能差距符号反转。因此，有效距离与较大的汉明距离未能带来增益这一结果相符，但它无法确定性能差距的大小；使用重采样残差时，甚至无法确定其符号。残差取决于输入（附录 I），而这些预测器都未对这一点建模。

(iv) 参考解码器。 P_{floor} 是给定分类器上的 LL 解码错误率，而非贝叶斯错误率；要估计训练后的分类器的噪声水平，需要真实后验概率，而目前尚无仅根据分类器输出稳定估计该概率的方法（Abe & Kumoi, 2026），因此无法根据该理论计算真实数据上的 e_{lower} 和 d^{*}。此外，该理论针对乘积解码器 (4)，而图 2(b) 使用求和解码；两者在 ± 0.005 范围内的一致性（第 9.1 节）是经验性的。

启示。在这些局限范围内，该理论给出了一条设计规则：最小距离必须增大，使得 $\kappa\, d_{\min}$ 超过 $\ln(M-1)$；超过饱和距离后，延长编码只会增加分类器数量。由于 $e_{\mathrm{lower}} \propto 1/\sigma^{2}$ 和 σ 取决于二元分类器的训练方式，饱和距离并非任务本身的属性。必要性部分为一种现象提供了可能的解释：编码 M 个类别、使用 $\lceil \log_2 M \rceil$ 对数编码的输出层，为何在未增加冗余时会损失准确率（Oda 等，2017）：对于 $d_{\min}=1$，只要二元分类器的噪声始终有一个大于零的下界，无论如何改进分类器，超额错误率都不会趋于零。

11 结论

我们将 Kumoi 等（2023）对噪声二元分类器的分析从成对错误概率扩展到分类错误率，并用最小汉明距离刻画了 ECOC 相对于 MAP 分类器的超额错误率。两个最可能类别的码字之间距离有界，会使超额错误率在 M 中始终远离零；最小距离若以快于 $\ln M$ 的速度增长，则会使其呈指数级趋于零（对于满足命题 5 条件的一般码）；而对于 nvR 码（$n \geq 2$），其精确指数为可靠性函数 E_{ECOC}，该函数在 $\Omega(\lambda)$ 上的最坏情况为 $\kappa_2(\lambda,\ \sigma)$。在相关高斯噪声下，将汉明距离替换为有效距离后，上述结论仍然成立；而类别级噪声会对有效距离形成界限（定理 2）。未来的研究方向包括哈达玛码及其他非 nvR 码的指数、以距离谱表示的一般码界限、包含前因子的有限长度分析，以及对训练后分类器有效距离的估计。

致谢

本研究得到日本学术振兴会科研费（资助编号：JP24K20833）的支持。

参考文献

Noriaki Abe 和 Gendo Kumoi。logit 噪声模型下贝叶斯最优校准的理论分析。机器学习研究汇刊，2026年。ISSN 2835-8856。URL https://openreview.net/forum?id=r8F3dqoopt。Erin L. Allwein、Robert E. Schapire 和 Yoram Singer。将多类别分类约简为二元分类：间隔分类器的统一方法。机器学习研究杂志，1:113–141，2000年。Jean-Yves Audibert 和 Alexandre B. Tsybakov。插件分类器的快速学习速率。统计年鉴，35(2):608–633，2007年。doi: 10.1214/009053606000001217。Peter L. Bartlett、Michael I. Jordan 和 Jon D. McAuliffe。凸性、分类与风险界。美国统计学会杂志，101(473):138–156，2006年。doi: 10.1198/016214505000000907。Alina Beygelzimer、Varsha Dani、Thomas P. Hayes、John Langford 和 Bianca Zadrozny。分类任务之间的误差限制约简。载于第*22*届国际机器学习会议（*ICML*）论文集，第49–56页，2005年。doi: 10.1145/1102351.1102358。

Alina Beygelzimer、John Langford 和 Pradeep Ravikumar。纠错锦标赛。载于算法学习理论（*ALT*），计算机科学讲义第5809卷，第247–262页。Springer，2009年。doi: 10.1007/978-3-642-04414-4_22。Thomas G. Dietterich 和 Ghulum Bakiri。通过纠错输出码解决多类学习问题。人工智能研究杂志，2:263–286，1995年。doi: 10.1613/jair.105。Sergio Escalera、Oriol Pujol 和 Petia Radeva。纠错输出码库。机器学习研究杂志，11:661–664，2010年。William Feller。概率论及其应用导论，第1卷。John Wiley & Sons，第3版，1968年。Robert G. Gallager。编码定理的简明推导及其若干应用。电气电子工程师学会信息论汇刊，11(1):3–18，1965年。doi: 10.1109/TIT.1965.1053730。Robert G. Gallager。信息论与可靠通信。John Wiley & Sons，1968年。Robert D. Gordon。自变量取大值时，面积与边界纵坐标之比（米尔斯比率）及正态概率积分的数值。数学统计年鉴，12(3):364–366，1941年。doi: 10.1214/aoms/1177731721。Shigeichi Hirasawa、Gendo Kumoi、Hideki Yagi、Manabu Kobayashi、Masayuki Goto 和 Hiroshige Inazumi。采用构造性编码的纠错输出码构造方法及其系统评估。载于*2022*年*IEEE*系统、人和控制论国际会议（*SMC*）论文集，第3053–3059页，2022年。doi: 10.1109/SMC53654.2022.9945163。Gareth James 和 Trevor Hastie。纠错编码方法与PICTs。计算与图形统计学杂志，7(3):377–387，1998年。doi: 10.1080/10618600.1998.10474782。Manabu Kobayashi、Gendo Kumoi、Hideki Yagi 和 Shigeichi Hirasawa。基于纠错输出码的多分类平均性能分析。载于*2024*年*IEEE*系统、人和控制论国际会议（*SMC*）论文集，第5373–5378页，2024年。doi: 10.1109/SMC54092.2024.10831957。Gendo Kumoi、Hideki Yagi、Manabu Kobayashi、Masayuki Goto 和 Shigeichi Hirasawa。多标签分类问题中ECOC方法优化研究（日文）。*IPSJ*数学建模及其应用汇刊（*TOM*），14(3):1–10，2021年。Gendo Kumoi、Hideki Yagi、Manabu Kobayashi 和 Shigeichi Hirasawa。汉明距离对采用估计二元分类器的ECOC性能的影响。载于*2022*年*IEEE*系统、人和控制论国际会议（*SMC*）论文集，第111–116页，2022年。doi: 10.1109/SMC53654.2022. 9945575。Gendo Kumoi、Hideki Yagi、Manabu Kobayashi、Masayuki Goto 和 Shigeichi Hirasawa。基于有噪声和无噪声二元分类器的纠错输出编码性能评估。国际神经系统杂志，33(2):2350004，2023年。doi: 10.1142/S0129065723500041。John Langford 和 Alina Beygelzimer。敏感纠错输出码。载于学习理论：第*18*届年度学习理论会议（*COLT*），计算机科学讲义第3559卷，第158–172页。Springer，2005年。doi: 10.1007/11503415_11。Vladimir I. Levenshtein。Hadamard矩阵在编码理论问题中的应用。控制论问题，5:125–136，1961年。F. Jessie MacWilliams 和 Neil J. A. Sloane。纠错码理论。North-Holland，阿姆斯特丹，1977年。

Hieu D. Nguyen、Mohammed Sarosh Khan、Nicholas Kaegi、Shen-Shyang Ho、Jonathan Moore、Logan Borys 和 Lucas Lavalva。使用纠错输出码的集成学习：新的分类错误界。载于电气电子工程师学会第 *33* 届国际人工智能工具会议（*ICTAI*）论文集，第 719–723 页，2021 年。doi：10.1109/ICTAI52525.2021.00114。

Yusuke Oda、Philip Arthur、Graham Neubig、Koichiro Yoshino 和 Satoshi Nakamura。通过二进制码预测实现神经机器翻译。载于第 *55* 届计算语言学协会年会（*ACL*）论文集，第 850–860 页，2017 年。doi：10.18653/v1/P17-1079。

Morris Plotkin。具有指定最小距离的二进制码。*IRE* 信息论汇刊，6(4)：445–450，1960 年。doi：10.1109/TIT.1960.1057584。

Harish G. Ramaswamy、Balaji Srinivasan Babu、Shivani Agarwal 和 Robert C. Williamson。《多类学习问题中基于输出码的学习算法的一致性》。载于第 *27* 届学习理论会议（*COLT*）论文集，机器学习研究论文集第 35 卷，第 885–902 页，2014 年。

Ryan Rifkin 和 Aldebaro Klautau。《为一对其余分类辩护》。机器学习研究杂志，第 5 卷：101–141，2004 年。

Anderson Rocha 和 Siome Klein Goldenstein。《从二元分类到多类别分类：拓展一对其余、一对一和基于纠错输出码的方法》。*IEEE* 神经网络与学习系统汇刊，25(2)：289–302，2014 年。doi：10.1109/TNNLS.2013.2274735。

Ambuj Tewari 和 Peter L. Bartlett。《多类分类方法的一致性》。机器学习研究杂志，第 8 卷：1007–1025，2007 年。

A. Viterbi 和 J. Omura。《数字通信与编码原理》。麦格劳-希尔，1979年。

A 引理 1 的证明（ η的上界）

由成对条件 $w_{lj} = w_{lj'}\,(l \neq 1,\ k)$， $f_{j'}$ 侧的和满足 $\alpha_{j'} = 1 - \theta_1 - \theta_k - \alpha_j$。代入该式，并将 η_j 展开为仅关于 α_j 的函数，得到

$$\eta_j(\alpha_j;\theta_1,\theta_k) = \frac{(1-\theta_1-\alpha_j)(\theta_k+\alpha_j)}{(\theta_1+\alpha_j)(1-\theta_k-\alpha_j)}, \qquad \alpha_j \in [0, 1-\theta_1-\theta_k]. \tag{60}$$

对其对数求导，并将各项两两合并，得

$$\frac{\partial \ln \eta_j}{\partial \alpha_j} = \frac{(\theta_1-\theta_k)(1-\theta_1-\theta_k-2\alpha_j)}{(\theta_1+\alpha_j)(\theta_k+\alpha_j)(1-\theta_1-\alpha_j)(1-\theta_k-\alpha_j)}. \tag{61}$$

分母在定义域上为正，因此对于 $\theta_1 > \theta_k$，导数在 $\alpha_j < (1-\theta_1-\theta_k)/2$ 时为正，在 $\alpha_j > (1-\theta_1-\theta_k)/2$ 时为负。因此， η_j是单峰函数，并在 $\alpha_j = (1-\theta_1-\theta_k)/2$ 处取得最大值；代入该值即可得到（26）。□

B 引理 4 的证明（ η的下界）

由式 (61)（附录 A）可知，对于 $\theta_1 > \theta_k$，式 (60) 中的函数 η_j 在 $[0, (1-\theta_1-\theta_k)/2]$ 上递增，在 $[(1-\theta_1-\theta_k)/2,\ 1-\theta_1-\theta_k]$ 上递减。因此，它在闭区间上的最小值在某个端点处取得。我们计算两个端点处的函数值。

在 $\alpha_j = 0$：

$$\eta_j(0;\theta_1,\theta_k) = \frac{(1-\theta_1)\theta_k}{\theta_1(1-\theta_k)}. \tag{62}$$

在 $\alpha_j = 1-\theta_1-\theta_k$：

$$\eta_j(1-\theta_1-\theta_k;\theta_1,\theta_k) = \frac{\theta_k(1-\theta_1)}{(1-\theta_k)\theta_1}. \tag{63}$$

这与 (62) 一致。由于 η_j 在定义域的两个端点处取值相同，因此最小值就是该端点处的值：

$$\eta_j(\alpha_j;\theta_1,\theta_k) \geq \frac{(1-\theta_1)\theta_k}{\theta_1(1-\theta_k)}. \tag{64}$$

□

C 引理 7 的证明（均匀指数与逐点指数之间的关系）

令 $h(\boldsymbol{\theta}) \triangleq \frac{\theta_1}{(1-\theta_2)\,(1-\theta_1)\theta_2}$，使得 $E_{\mathrm{ECOC}} = \frac{1}{8\sigma^2}\ln^2 h$，且 $\ln\, h > 0$，因为 $\theta_1 > \theta_2$。作变量替换，令 $s = \theta_1+\theta_2 \leq 1$ 和 $\delta = \theta_1-\theta_2 \geq \lambda$，

$$\frac{\partial \ln h}{\partial s} = 2\delta\left[\frac{1}{(2-s)^2-\delta^2} - \frac{1}{s^2-\delta^2}\right] \leq 0, \tag{65}$$

其中不等式由 $s \leq 1 \leq 2-s$ 得出。因此，$\ln h$ 关于 s 单调递减，并在 $s=1$ 处取得最小值 $\ln\, h = 2\ln \frac{1+\delta}{1-\delta}$。此外，$\ln \frac{1+\delta}{1-\delta}$ 关于 δ 单调递增，因此其最小值在 $\delta=\lambda$ 处取得，且

$$\begin{aligned} E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sigma) &\geq \frac{1}{8\sigma^2}\left(2\ln\frac{1+\lambda}{1-\lambda}\right)^2 \\ &= \frac{1}{2\sigma^2}\ln^2\!\left(\frac{1+\lambda}{1-\lambda}\right) = \kappa_2(\lambda,\sigma). \end{aligned} \tag{66}$$

□

D 推论 4 的证明（穷举码的渐近评估）

推论 4（穷举码的渐近评估）。对于 $\boldsymbol{\theta} \in \Omega(\lambda)$，对任意 $\zeta > 0$ 以及充分大的 M：

上界：

$$\Delta(\boldsymbol{\theta}) \leq \exp\left\{-N_{\mathrm{exh}}\left(\frac{\kappa_2(\lambda,\sigma)}{2} - \zeta\right)\right\}. \tag{67}$$

下界：

$$\Delta(\boldsymbol{\theta}) \geq \Theta\left(\frac{1}{\sqrt{N_{\mathrm{exh}}}}\right)\exp\{-N_{\mathrm{exh}}\,\delta_{\mathrm{exh}}\,e_{\mathrm{exh}}(\boldsymbol{\theta})\}, \tag{68}$$

其中 $N_{\mathrm{exh}} = 2^{M-1}-1$，$\delta_{\mathrm{exh}} = d_{\mathrm{exh}}/N_{\mathrm{exh}} \in (1/2,1]$，且 $e_{\mathrm{exh}}(\boldsymbol{\theta}) \triangleq \frac{1}{8\sigma^2}\ln^2 \frac{\theta_1}{(1-\theta_2)\,(1-\theta_1)\theta_2}$。

上界。对每个块应用命题4并求和，得到 $\mu_{1k} \leq d_{\mathrm{exh}} \ln \frac{1-\theta_1+\theta_k}{1+\theta_1-\theta_k}$；与推论 1 中一样，我们得到 $\Delta(\boldsymbol{\theta}) \leq (M-1)\exp(-d_{\mathrm{exh}}\,\kappa_2(\lambda,\sigma))$。由 $d_{\mathrm{exh}} = N_{\mathrm{exh}}\delta_{\mathrm{exh}}$，$\delta_{\mathrm{exh}} > 1/2$，以及 $M-1 = \log_2(N_{\mathrm{exh}}+1) \leq e^{\zeta N_{\mathrm{exh}}}$（对充分大的 M），我们得到（67）。

下界。将命题 6 应用于每个块（对于 1vR 块，该命题取等号，见第 7.1 节）并求和，$\mu_{1k} \geq \frac{d_{\mathrm{exh}}}{2}\ln(\frac{1-\theta_1}{\theta_k})$；根据与命题 7 相同的论证，我们得到 $\Delta(\boldsymbol{\theta}) \geq \frac{\lambda}{2\sqrt{2\pi}\,\rho^{\mathrm{U}}}\exp(-(\rho^{\mathrm{U}})^2/2)$，其中 $(\rho^{\mathrm{U}})^2/2 = d_{\mathrm{exh}}\,e_{\mathrm{exh}}(\boldsymbol{\theta})$。由于 $\rho^{\mathrm{U}} = \sqrt{2d_{\mathrm{exh}}\,e_{\mathrm{exh}}(\boldsymbol{\theta})} = \Theta(\sqrt{N_{\mathrm{exh}}})$，前因子为 $\Theta(1/\sqrt{N_{\mathrm{exh}}})$。□

E 引理 6 的证明（μ_{12}的渐近精确值）

步骤 *1*： μ_{12} 的精确表达式。

分类器 f_j 分隔 c_1 和 c_2 时，其正侧群组形式为 $\{c_1\} \cup \mathcal{S}$ ($|\mathcal{S}| = n-1, \mathcal{S} \subseteq \{3,\dots,M\}$)。令 $\alpha_{\mathcal{S}} \triangleq \sum_{i\in \mathcal{S}} \theta_i$,

$$\mu_{12} = \sum_{\mathcal{S}\subseteq\{3,\dots,M\},|\mathcal{S}|=n-1} \ln \frac{(1-\theta_1-\alpha_{\mathcal{S}})(\theta_2+\alpha_{\mathcal{S}})}{(\theta_1+\alpha_{\mathcal{S}})(1-\theta_2-\alpha_{\mathcal{S}})}. \tag{69}$$

此表达式假设 $M > 2n$。对于 $n = M/2$，每一对在求和中出现两次，一次以 $\mathcal{S}$ 的形式出现，另一次以其在 $\{3,\dots,M\}$ 中的补集形式出现，且两项相等，因此右侧必须乘以 $1/2$；这是 $M = 4$ 在表3中采用的值。在两种情况下，不同项的数量均为 $d(n,M)/2$，因此 $\mu_{12}/(d(n,M)/2)$ 是对所有满足 $|\mathcal{S}| = n-1$ 的集合 $\mathcal{S} \subseteq \{3,\dots,M\}$ 上的 $\gamma(\alpha_{\mathcal{S}})$ 的平均值，其中

$$\gamma(\alpha) \triangleq \ln \frac{(1-\theta_1-\alpha)(\theta_2+\alpha)}{(\theta_1+\alpha)(1-\theta_2-\alpha)}, \qquad \alpha \in [0, 1-\theta_1-\theta_2]. \tag{70}$$

步骤 *2*：对 γ 的双侧界。

由于 $\gamma = \ln \eta_j$，其中 η_j 如式 (60) 所示（$\theta_k = \theta_2$），由引理 4 可得，$\gamma(\alpha)$ 在整个定义域上 $\geq \gamma(0)$；根据 (61)，γ 先严格递增，随后在两个取值相同的端点之间严格递减，因此对于 $0 < \alpha < 1-\theta_1-\theta_2$，$\gamma(\alpha) > \gamma(0)$。对于上界，将 $\gamma(\alpha) - \gamma(0)$ 写成 $\ln \frac{1-\theta_1-\alpha}{1-\theta_1} \leq -\frac{\alpha}{1-\theta_1}$、$\ln \frac{\theta_2+\alpha}{\theta_2} \leq \frac{\alpha}{\theta_2}$、$-\ln \frac{\theta_1+\alpha}{\theta_1} \leq -\frac{\alpha}{\theta_1+\alpha} \leq -\frac{\alpha}{1-\theta_2}$ 和 $\ln \frac{1-\theta_2}{1-\theta_2-\alpha} \leq \frac{\alpha}{1-\theta_2-\alpha} \leq \frac{\alpha}{\theta_1}$ 之和，其中我们使用了 $\ln x \leq x-1$（对于第三项，则使用 $\ln x \geq 1-1/x$），以及 $\theta_1+\alpha \leq 1-\theta_2$ 和 $1-\theta_2-\alpha \geq \theta_1$。因此，令 $c' \triangleq \frac{1}{\theta_1} + \frac{1}{\theta_2} - \frac{1}{1-\theta_1} - \frac{1}{1-\theta_2} \geq 0$,

$$0 \leq \gamma(\alpha) - \gamma(0) \leq c'\alpha, \qquad \alpha \in [0, 1-\theta_1-\theta_2]. \tag{71}$$

步骤 *3*：求值 $\bar{\alpha}_{\mathcal{S}}$。

每个 $i \in \{3,\dots,M\}$ 都包含在 $\binom{M-3}{n-2}$ 个集合 $\mathcal{S}$的 $\binom{M-2}{n-1}$ 中，因此

$$\bar{\alpha}_{\mathcal{S}} = \frac{n-1}{M-2}(1-\theta_1-\theta_2). \tag{72}$$

这是一个只用到 $\sum_{i\geq 3}\theta_i = 1-\theta_1-\theta_2$的恒等式。

步骤 *4*：论断的推导。

定义有限 M 时与渐近目标值之间的误差为

$$r(M) \triangleq \frac{\mu_{12}}{d(n,M)/2} - \ln \frac{(1-\theta_1)\theta_2}{\theta_1(1-\theta_2)} = \overline{\gamma(\alpha_{\mathcal{S}}) - \gamma(0)}. \tag{73}$$

对$\mathcal{S}$ 上的 (71) 取平均，并使用 (72),

$$0 \leq r(M) \leq c'\,\bar{\alpha}_{\mathcal{S}} = \frac{\beta(\theta_1,\theta_2,n)}{M-2} \tag{74}$$

对于每个 $M \geq 2n$，这就是（53），并且意味着（52）。如果 $\theta_i > 0$ 对于所有 $i \geq 3$均成立，那么 $0 < \alpha_{\mathcal{S}} < 1-\theta_1-\theta_2$ 对于每个 $\mathcal{S}$均成立（因为 $n-1 < M-2$），因此根据步骤 2，每一项都严格为正，并且 $r(M) > 0$ 对于每个 $M \geq 2n$均成立；也就是说，在有限 M下，$\mu_{12}/(d/2)$ 位于渐近值的较大一侧（绝对值较小）。因此，$\rho_{12} < \rho_{12}^{\mathrm{U}}$以及 $\rho_{12}^2/(2d(n,M))$ 在每个有限 M下均成立：修正项降低了成对误差的高斯指数。

$r(M)$ 的首项由围绕 $\alpha = 0$ 的泰勒展开给出。根据（61），

$$\gamma'(0) = -\frac{1}{1-\theta_1} + \frac{1}{\theta_2} - \frac{1}{\theta_1} + \frac{1}{1-\theta_2} = \frac{(\theta_1-\theta_2)(1-\theta_1-\theta_2)}{\theta_1\theta_2(1-\theta_1)(1-\theta_2)}, \tag{75}$$

当且仅当 $1-\theta_1-\theta_2>0$（给定 $\theta_1>\theta_2$）时，该式为正；这在引理的设定下成立。在 $\theta_i=O(1/M)$（$i\ge 3$）条件下，我们有 $\alpha_{\mathcal{S}}^2=O(1/M^2)$，因此

$$r(M)=\frac{\gamma'(0)(n-1)(1-\theta_1-\theta_2)}{M-2}+O\!\left(\frac{1}{M^2}\right), \tag{76}$$

并且 $r(M)\times M$ 收敛于常数 $\gamma'(0)(n-1)(1-\theta_1-\theta_2)$，由于 $c'-\gamma'(0)=2/\theta_1-2/(1-\theta_2)\ge 0$，该常数至多为 $\beta(\theta_1,\ \theta_2,\ n)$；这一点可通过数值计算加以确认（附录 H.1）。

步骤 *5*：对可靠性函数的贡献。

由于 $r(M)\to 0$，$\rho_{12}^2/(2d(n,\ M))\to \frac{1}{8\sigma^2}\ln^2\frac{\theta_1}{\theta_2}=e_{\text{lower}}$。 □

F 命题 8 的证明（可靠性函数）

令 $e_{\text{lower}}\triangleq \frac{1}{8\sigma^2}\ln^2\!\left(\frac{\theta_1}{\theta_2}\cdot\frac{}{(1-\theta_2)\,(1-\theta_1)}\right)$。

竞争码字 $k\ge 3$ 可忽略。令 $d=d(n,M)$，并令 c_0 为满足 $\theta_i\le c_0/M$ 的常数，其中 $i\ge 3$（定义 6）。对于 $M>2n$，按照式（69），以 k 替代 2，$\mu_{1k}=\sum_{\mathcal{S}}\gamma_k(\alpha_{\mathcal{S}})$，其中 $\mathcal{S}$ 遍历集合 $d/2$ 子集 $\{2,\dots,M\}\setminus\{k\}$，其大小为 $n-1$，$\gamma_k(\alpha)\triangleq \ln\frac{(1-\theta_1-\alpha)(\theta_k+\alpha)}{(\theta_1+\alpha)(1-\theta_k-\alpha)}$，并且根据引理 1，每一项均为负。如果 $\theta_k+\alpha_{\mathcal{S}}\le 1/2$，则 $-\gamma_k(\alpha_{\mathcal{S}})\ge \ln\frac{\theta_1}{(\theta_k+\alpha_{\mathcal{S}})}$。如果 $2\notin\mathcal{S}$，则 $\theta_k+\alpha_{\mathcal{S}}\le nc_0/M$，包含 2 的集合所占比例为 $(n-1)/(M-2)$。因此

$$\min_{k\ge 3}\frac{-\mu_{1k}}{d}\ \ge\ \frac{1}{2}\Big(1-\frac{n-1}{M-2}\Big)\Big(\ln M-\ln\frac{2nc_0}{\theta_1}\Big)\ \longrightarrow\ \infty, \tag{77}$$

这对每个 $M>2n$ 都成立（当右侧为负时，该结论显然成立，因为左侧非负），而 $-\mu_{12}/d$ 根据引理 6 收敛；由于 $\rho_{1k}=-\mu_{1k}/(\sigma\sqrt{d})$，对于所有充分大的 M，我们有 $\rho_{\min}=\rho_{12}$。

上界和下界。由并集界和 Chernoff界（引理 2 的上界）可得 $P(E|\boldsymbol{\theta})$ 的一个上界，

$$P(E|\boldsymbol{\theta})\le (M-1)\exp\!\left(-\frac{\rho_{\min}^2}{2}\right), \tag{78}$$

由此得到 $-\ln P(E|\boldsymbol{\theta})$ 的一个下界。根据引理 6（附录 E），对于所有充分大的 M，我们有 $\rho_{\min}=\rho_{12}$、$\rho_{12}^2/(2d(n,M))\to e_{\text{lower}}$；并且对于任意 $\zeta>0$，都有 $(M-1)\le e^{\zeta d(n,M)}$（因为 $\ln(M-1)=o(d(n,M))$），因此

$$\liminf_{M\to\infty}\frac{-\ln P(E|\boldsymbol{\theta})}{d(n,M)}\ge e_{\text{lower}}. \tag{79}$$

反过来，由 $\{R_{12}\le 1\}\subseteq E$ 以及引理 2 的下界（当 $\rho_{12}\ge 1$ 时），我们得到 $P(E|\boldsymbol{\theta})$ 的一个下界，

$$P(E|\boldsymbol{\theta})\ge \frac{1}{2\sqrt{2\pi}\rho_{12}}\exp\!\left(-\frac{\rho_{12}^2}{2}\right), \tag{80}$$

这给出了关于 $-\ln P(E|\boldsymbol{\theta})$ 的一个上界。由于 $\ln\rho_{12}=O(\ln d(n,\ M))$ 关于 $d(n,\ M)$ 是次线性的，

$$\limsup_{M\to\infty}\frac{-\ln P(E|\boldsymbol{\theta})}{d(n,M)}\le e_{\text{lower}}. \tag{81}$$

由于下极限 $\le$ 上极限总是成立，上述两个界迫使下极限 $=$ 上极限 $=e_{\text{lower}}$，因此极限存在，并且 $E_{\text{ECOC}}(\boldsymbol{\theta},\ \sigma)=e_{\text{lower}}$。 □

G 定理 2 的证明（相关高斯分布噪声）

第 *1* 部分，充分性和必要性。对于一般的协方差，(7) 和 (14) 给出 $P(R_{1k}\le 1|\boldsymbol{\theta}) = Q(\rho_{1k})$，其中

$$\rho_{1k} \triangleq \frac{-\mu_{1k}}{\sqrt{\boldsymbol{s}_k^\top \Sigma \boldsymbol{s}_k}} = \frac{-\mu_{1k}}{d_H(c_1,c_k)}\sqrt{d_{\mathrm{eff}}(c_1,c_k)}, \tag{82}$$

它替代了 (9)。命题 2 中的噪声只通过 $P(E|\boldsymbol{\theta})$ 起作用，而命题 3、4 和 6 以及界 (41) 仅与 μ_{1k} 有关。由并集界和引理 2，$P(E|\boldsymbol{\theta}) \le \sum_{k\neq 1}\exp(-\rho_{1k}^2/2)$。根据 (24)，在命题 5 的条件下，$-\mu_{1k}/d_H(c_1,c_k) \ge \ln\frac{\theta_1+\omega}{1-\theta_1-\omega}(n_{\min})$；根据 (27)，~~对于 n vR 码~~，$-\mu_{1k}/d(n,M) \ge \ln\frac{1+\lambda}{1-\lambda}$；两种情况下都有 $\rho_{1k}^2/2 \ge \bar{\kappa}\, d_{\mathrm{eff}}(c_1,c_k)$，由此得到 (59) 中的上界。对于下界，$\{R_{12}\le 1\}\subseteq E$ 给出 $P(E|\boldsymbol{\theta}) \ge Q(\rho_{12})$，而 (41) 给出 $-\mu_{12} \le d_H(c_1,c_2)\,B_2(\boldsymbol{\theta})$，因此 $\rho_{12} \le \sqrt{d_{\mathrm{eff}}(c_1,c_2)}\,B_2(\boldsymbol{\theta})$；由于 Q 单调递减，结论由命题 2 得出。

第一部分，紧性。令 $d = d(n,M)$，固定 $n\ge 2$，取定义 6 中的族，并令 $V \mathrm{bound}\sigma_{\mathrm{eff}}^2(1,k)$。根据（77）（附录 F），该式仅涉及 μ_{1k}，$-\mu_{1k}/d \to \infty$ 关于 $k\ge 3$均匀成立，因此 $\rho_{1k}^2/(2d) \ge (\mu_{1k}/d)^2/(2V) \to \infty$ 关于 $k\ge 3$均匀成立。由引理 6，$\rho_{12}^2/(2d) = (\mu_{12}/d)^2/(2\sigma_{\mathrm{eff}}^2(1,2)) \to E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sqrt{v})$。因此，对于所有充分大的 M，$\rho_{\min} = \rho_{12}$ 成立，并且附录 E 的论证适用于将 $\sigma\sqrt{d}$ 替换为 $(\boldsymbol{s}_2^\top\Sigma\boldsymbol{s}_2)^{1/2}$ 的情形：$Q(\rho_{12}) \le P(E|\boldsymbol{\theta}) \le (M-1)\exp(-\rho_{12}^2/2)$，且 $\ln(M-1) = o(d)$ 给出 $-\ln P(E|\boldsymbol{\theta})/d \to E_{\mathrm{ECOC}}(\boldsymbol{\theta},\sqrt{v})$，而命题 2 按照推论 3 将极限转移到 $\Delta(\boldsymbol{\theta})$。

第 *2* 部分。由于 $\mathrm{Cov}(\varepsilon_j,\ \varepsilon_{j'}) = \sigma_0^2\mathbf{1}_{\{j=j'\}} + \sigma_{\mathrm{c}}^2\sum_i w_{ij}w_{ij'} + \sigma_{\mathrm{z}}^2$，我们有 $\Sigma = \sigma_0^2 I + \sigma_{\mathrm{c}}^2 W^\top W + \sigma_{\mathrm{z}}^2\mathbf{1}\mathbf{1}^\top$。令 $\boldsymbol{e}_i$ 为 $\mathbb{R}^M$ 的第 i 个单位向量，则根据柯西–施瓦茨不等式，$d_H(c_1,\ c_k) = \|\boldsymbol{s}_k\|^2 = (\boldsymbol{e}_k-\boldsymbol{e}_1)^\top W\boldsymbol{s}_k \le \sqrt{2}\,\|W\boldsymbol{s}_k\|$，因此 $\boldsymbol{s}_k^\top\Sigma\boldsymbol{s}_k \ge \sigma_{\mathrm{c}}^2\|W\boldsymbol{s}_k\|^2 \ge \sigma_{\mathrm{c}}^2 d_H(c_1,\ c_k)^2/2$，也就是说，$d_{\mathrm{eff}}(c_1,\ c_k) \le 2/\sigma_{\mathrm{c}}^2$；$\Delta(\boldsymbol{\theta})$ 的下界由第 1 部分得出。令 W 为距离 d 且行重量为 ℓ 的等距集合。则 $WW^\top = (d/2)I + (\ell - d/2)\mathbf{1}\mathbf{1}^\top$，而式 (13) 中 L_i 的噪声部分为

$$\sum_j w_{ij}\varepsilon_j = \sum_j w_{ij}\xi_j + \frac{d}{2}\chi_i + \Big(\ell - \frac{d}{2}\Big)\sum_l \chi_l + \ell z. \tag{83}$$

最后两项对所有 i 都相同。前两项 $\sigma_0^2 WW^\top + \sigma_{\mathrm{c}}^2(d/2)^2 I$ 的协方差与 $(\sigma_0^2 + \sigma_{\mathrm{c}}^2 d/2)\ WW^\top$ 不同；后者是假设1下 $\sigma^2 = \sigma_0^2 + \sigma_{\mathrm{c}}^2 d/2$ 时的协方差，两者相差 $\mathbf{1}\mathbf{1}^\top$ 的倍数，而该项在每个差值 $L_i - L_k$ 中都会抵消。两个模型中的得分差具有相同的均值和协方差，且均服从联合高斯分布，因此 i 具有相同的分布。如果 $\sigma_{\mathrm{c}} = 0$，则只剩下 ℓz 这一项，相等行权重便已足够。□

H 数值验证

本附录将对截至第6节的理论结果进行数值验证。此处的设置对应于 L0 设置（第 9.1 节），它完全满足理论假设（对数几率空间中方差 σ^2 已知的独立同分布高斯噪声），在此设置下，可以对可靠性函数的值 e_{lower} 进行定量验证。我们取 $n=2$（2vR编码）、$\theta_1 = 0.5$、$\theta_2 = 0.25$（因此 $\lambda = \theta_1 - \theta_2 = 0.25$）、$\sigma = 1.0$，并将其余类别的后验概率均匀分配为 $\theta_i = (1-\theta_1-\theta_2)/(M-2)$（$i\ge 3$）。对于每个 M，我们使用 500,000 次试验进行蒙特卡罗估计。

H.1 引理验证：$\mu_{12}/(d/2)$ 的渐近行为

引理 6 论断，当 $M\to\infty$（n 固定）时，

$$\frac{\mu_{12}}{d(n,M)/2} \longrightarrow \ln\frac{(1-\theta_1)\theta_2}{\theta_1(1-\theta_2)} \tag{84}$$

并保证误差为 $O(1/M)$。表3 将左侧的精确数值与渐近目标值进行了比较。

表3：引理 6 的数值验证（$n=2$, $\theta_1=0.5$, $\theta_2=0.25$, $\sigma=1.0$）。

M	$d(2,M)$	$\mu_{12}/(d/2)$	渐近目标	误差 $r(M)$	$\lvert r(M)\rvert\times M$
4	2	-1.02165	-1.09861	$+0.07696$	0.308
6	8	-1.03977	-1.09861	$+0.05884$	0.353
10	16	-1.06343	-1.09861	$+0.03518$	0.352
20	36	-1.08142	-1.09861	$+0.01719$	0.344
50	96	-1.09186	-1.09861	$+0.00675$	0.338
100	196	-1.09526	-1.09861	$+0.00336$	0.336
500	996	-1.09794	-1.09861	$+0.00067$	0.334

表 4：　可靠性函数的数值验证（$n=2$, $\theta_1=0.5$, $\theta_2=0.25$, $\sigma=1.0$）。

M	d	$P(E)_{\mathrm{MC}}$	$\frac{-\ln P(E)_{\mathrm{MC}}}{d}$	e_{lower}	κ_2（均匀）	$-\ln Q(\rho^U)/d$
4	2	0.3384	0.5418	0.1509	0.1305	0.7602
6	8	0.0728	0.3276	0.1509	0.1305	0.3514
8	12	0.0341	0.2816	0.1509	0.1305	0.2964
10	16	0.0166	0.2563	0.1509	0.1305	0.2668
15	26	0.0030	0.2239	0.1509	0.1305	0.2297
20	36	0.0006	0.2061	0.1509	0.1305	0.2117

$-\ln Q(\rho^U)/d$：引理 5 的下界 $Q(\rho^{\mathrm{U}}_{12})$ 的指数。由于 $(\rho^{\mathrm{U}}_{12})^2/2 = d\,e_{\mathrm{lower}}$ 恰好成立，它仅通过 Q 的前因子超过 e_{lower}，并在 $d\to\infty$ 时收敛至 e_{lower}。在 $M=4$ 时，2vR 编码具有 $n=M/2$，因此根据（2），$d(2,4) = \binom{2}{1} = 2$。

$|r(M)|\times M$在 $M=4$至 $M=500$范围内收敛至约 0.333（$=1/3$），证实误差为 $O(1/M)$；此外，每个条目还满足引理中的显式界 $0<r(M)\le\beta/(M-2)$，此处为 $\beta=2/3$。另外，误差 $r(M)$ 为正（根据引理 6，对任意 M 均成立），因此有限 M 下的 $\mu_{12}/(d/2)$ 大于渐近目标值，即其绝对值更小。因此，在每个有限 M 下，$\rho_{12}<\rho^{\mathrm{U}}_{12}$和 $\rho^2_{12}/(2d(n,M)) < e_{\mathrm{lower}}$：引理 6 中的修正项将成对误差 $Q(\rho_{12})$ 的高斯指数降低至 e_{lower} 以下，并将 $Q(\rho_{12})$ 提高至 $Q(\rho^{\mathrm{U}}_{12})$以上。例如，在 $M=20$时，$\rho^2_{12}/(2d)=0.1462<e_{\mathrm{lower}}=0.1509$ 以及$Q(\rho_{12})=5.89\times10^{-4} > Q(\rho^{\mathrm{U}}_{12})=4.91\times10^{-4}$。

H.2 主要定理的验证：可靠性函数的收敛

命题 8 声称，对于一个 nvR 编码，其中 $n\ge2$（此处为 2vR编码），

$$\lim_{M\to\infty}\frac{-\ln P(E|\boldsymbol{\theta})}{d(n,M)} \;=\; e_{\mathrm{lower}} \;=\; \frac{1}{8\sigma^2}\ln^2\!\frac{\theta_1(1-\theta_2)}{(1-\theta_1)\theta_2}. \tag{85}$$

在此设定下，$e_{\mathrm{lower}}=0.15087$ 比在 $\Omega(\lambda)$ 上的均匀最坏情况指数 $\kappa_2(\lambda,\sigma)$ 大约高 15.6%（见引理 7）。

表 4 将指数 $-\ln P(E|\boldsymbol{\theta})_{\mathrm{MC}}/d$ 的蒙特卡洛估计值与两个公式进行了比较。

从表中可以得到以下观察结果。(1) 随着 M 增加，蒙特卡洛指数单调下降，并趋近于 $e_{\mathrm{lower}}=0.15087$。(2) 在有限 M 下，蒙特卡洛指数大于 e_{lower}。这种超出并非由引理 6 的修正项 $r(M)$造成的，因为该修正项的作用方向相反：$r(M)>0$ 使 $|\mu_{12}|$变小，因而使 ρ_{12}变小，所以 $\rho^2_{12}/(2d) < e_{\mathrm{lower}}$（见上一小节）。这种超出是由 Q的次指数前因子造成的。根据米尔斯比率界 $\phi(x)\,x/(x^2+1)\le Q(x)\le\phi(x)/x (x>0,\ \phi$ 标准正态密度），$-\ln Q(\rho)/d=\rho^2/(2d)+\ln(\sqrt{2\pi}\,\rho)/d+O(1/(\rho^2 d))$，并且对于 $n=2$，项 $\ln(\sqrt{2\pi}\,\rho_{12})/d=\Theta(\ln d/d)$ 超过性能差距 $e_{\mathrm{lower}}-\rho^2_{12}/(2d)=O(1/M)=O(1/d)$。事实上，$-\ln Q(\rho_{12})/d=0.3311$，当 $M=6$为 10 和 20 时分别为 0.2557 和 0.2066，与蒙特卡洛指数 0.3276、0.2563 和 0.2061 高度一致，其差异处于蒙特卡罗误差范围内。(3) 一致指数 $\kappa_2=0.13047$在每个 M处都低于蒙特卡洛指数，这与引理 7 的关系 $E_{\mathrm{ECOC}}\ge\kappa_2$ 一致。

表5：验证上界和下界的一致性（$n = 2$、 $\theta_1 = 0.5$、 $\theta_2 = 0.25$、 $\sigma = 1.0$，500,000 次试验）。

M	d	$P(E)_{\mathrm{MC}}$	下界 $Q(\rho^U)$	上界 $(M-1)\,Q(\rho^L)$	一致性
4	2	0.3384	0.2186	0.7051	✓
5	6	0.1165	0.0892	0.4217	✓
6	8	0.0728	0.0601	0.3713	✓
8	12	0.0341	0.0285	0.2688	✓
10	16	0.0166	0.0140	0.1846	✓
15	26	0.0030	0.0025	0.0644	✓
18	32	0.0011	0.0009	0.0328	✓
20	36	0.0006	0.0005	0.0207	✓

H.3 上界和下界一致性的验证

我们检验以下不等式

$$Q(\rho_{12}^{\mathrm{U}}) \;\le\; P(E|\boldsymbol{\theta}) \;\le\; (M-1)\,Q(\rho^{\mathrm{L}}). \tag{86}$$

左侧不等式是引理 5 的下界。右侧不等式由并集界和$\rho_{1k} \ge \rho^{\mathrm{L}}$（式（32））推出；它比引理 5 的 Chernoff形式（$M-1$）exp（$-d\,\kappa_2$）更紧，而后者是通过 $Q(x) \le e^{-x^2/2}$由前者得到的。表5将上界和下界与蒙特卡洛估计进行了比较。

对于所有已验证的 M（4 $\le M \le 20$），蒙特卡罗估计介于上界和下界（✓）之间。

随着 M 增大，两个界之间的绝对差距缩小，但它们的比值从 $M = 4$ 处的 3.2 增长到 $M = 20$ 处的 42。同一范围内，归一化差距（$\ln[(M-1)Q(\rho^{\mathrm{L}})] - \ln Q(\rho_{12}^{\mathrm{U}})$）从 0.59 降至 0.10，并趋于 $e_{\mathrm{lower}} - \kappa_2(\lambda,\sigma)$ $= 0.0204 > 0$，而非趋于零。两个界都不涉及 $r(M)$：下界使用 ρ_{12}^{U}，即 μ_{12} 的渐近值；上界使用均匀 ρ^{L}。两个界的指数分别收敛于 e_{lower} 和 $\kappa_2(\lambda,\sigma)$，因此，除非 $\theta_1+\theta_2 = 1$ 和 $\theta_1-\theta_2 = \lambda$（引理 7 的等号情形），引理 5 无法将 $P(E|\boldsymbol{\theta})$ 确定到一个次指数因子以内。精确指数 e_{lower} 可由上界（78）在命题 8 中得到，该上界的指数根据引理 6 收敛于 e_{lower} 。

I 实验细节

列子集（图 2(b)）。对于每个数据集，列池在不同随机种子之间保持固定：它由 CIFAR-10 的 2vR 和 3 vR 编码的所有 $28 + 56 = 84$ 列组成（$M = 8$），CIFAR-100 的全部 $66 + 220$ =286 列组成（$M = 12$），以及 ImageNet-1k 的 2vR 编码的 120 列和从 3vR 编码中随机选取的 200 列（总计 320 列）组成（$M = 16$）。对于 $K \in${3, 4, 5, 6, 8, 10, 14, 18, 24, 32, 44, 60, 80, 120, 180, 240, 300}（不超过列池大小），并且每个 K 最多进行 8 次随机抽取（含有重复码字的表将被丢弃），从列池中抽取 K 列，训练二元分类器，并记录所抽取表的最小距离及其求和解码错误率。根据最小距离对记录进行分箱，分箱边界为 0.5、1.5、2.5、3.5、4.5、6.5、8.5、11.5、15.5、20.5、27.5、37.5、50.5、67.5、90.5、120.5；图 2(b)绘制每个分箱中的平均距离以及 Δ_{real} 的平均值。ImageNet-1k 图像取自训练集的一个子集，且每个随机种子都会重新抽取类别。

解码规则和参考解码器。 在 CIFAR-100 上（ResNet-50 特征，经主成分分析降至 100 个维度，$M \in$ {4、 6、 8、 10 、 12 、 16}，25 个随机种子），在相同分类器上，2vR 编码的求和解码与 LL 解码错误率之差对于 $M = 6$–16 在统计上等效于 0（显著性水平为 0。05、等效性界值为 0。005 的 TOST等效性检验；该差异在实际意义上可忽略；最大值为$p = 1.2 \times 10^{-10}$），而对于 $M = 4$，每个随机种子下的差异都恰好为 0，因为两个解码器作出相同的决策。在相同的 150 个条件下，使用相同特征训练的多类逻辑回归在 136 个条件下的错误率高于对 2vR分类器进行 LL解码的错误率，在 8 个条件下相等，在 6 个条件下更低（平均差异为 0。011）。对于 $M = 4$ ，该实现将每个 2vR划分及其补集一同存储

（$n = M/2$）；由于使用互补标签训练的逻辑回归会输出 $1 - p_j$，因此这种重复不会改变任一解码器的决策（在求解器容差范围内）。

3vR 与 2vR 的比较。 第 9.2 节中的比较使用相同的 CIFAR-100 实验运行。比较范围限定为$M \in \{8,\ 10,\ 12\}$，这些取值满足 $d(3,\ M) > d(2,\ M)$：对于 $M = 6$，3vR 编码具有 $n = M/2$，并且根据式 (2)，$d(3,\ 6) = \binom{4}{2} = 6 < d(2,\ 6) = 8$；而对于 $M = 4$，3vR编码不符合定义 1。求和解码下，按随机种子匹配的差值 $\hat{}\, P_{\mathrm{err}}^{3\mathrm{vR}} - \hat{}\, P^{2\mathrm{vRerr}}$ 的均值为 $+0.0020$（Wilcoxon符号秩检验 $p = 1 \times 10^{-5}$，75 对）。

2vR 和 3vR 的有效距离（第 10(iii) 节）。教师模型是一个包含 100 个类别的多层感知机，其输入为标准化的 ResNet-50 特征，经主成分分析降至 256 维；该模型在 45,000 张 CIFAR-100训练图像上训练，并在剩余的 5,000 张图像上进行提前停止和温度校准（$T = 1.087$）；其测试准确率为 0.709。对于每个随机种子，以及与上述相同的 M 类别、图像、PCA-100特征和逻辑回归，标签从重新归一化到 M 类别的教师模型后验分布中重新抽样，$\boldsymbol{\theta}^*(\boldsymbol{x})$；对于 $M = 8$，10，12，其贝叶斯错误率分别为 0.065、0.084 和 0.103。由于已知 $\boldsymbol{\theta}^*$，我们报告期望错误率 $E[1 - \theta_i^*]$，不受标签抽样噪声影响；并报告对数优势残差 $r_j = \ln \frac{p_j}{1-p_j} - \ln \frac{f_j^*(1|\boldsymbol{\theta}^*)}{1-f_j^*(1|\boldsymbol{\theta}^*)}$。$\hat{}\,\Sigma$ 是 $(r_j)_j$ 在每类 100 张未用于训练的训练图像上的协方差；预测结果通过蒙特卡洛方法在测试集上评估（每张测试图像抽取 256 次）。按随机种子匹配的 3vR−2vR超额误差差值，对于 $M = 8$、10、12分别为 $+0.0003$、$+0.0014$ 和 $+0.0018$（每种情况 25 个随机种子）。残差与输入并非相互独立：将次优类别对的噪声对其间隔进行回归，得到的斜率为 -0.29 至 -0.42；每个随机种子下，3vR的斜率绝对值均更大，且其超额峰度为 0.75–1.16。所有预测器都将残差视为与输入独立，这很可能是幅度不匹配的来源。第 10(iii) 节中置信度较低的教师模型将温度加倍。

缩放 M（表2）。 我们选择 $M \in \{16、32、64、128、256\}$ 个 ImageNet-1k 类别，使用逻辑回归，在 ImageNet-1k 训练集的每个类别中抽取 200 张训练图像来训练 2vR 和 logM 编码，并在验证集上进行测试。所用特征为 ResNet-50（在 ImageNet1K_V2 上预训练）的特征，经主成分分析降至 128 维（$S = 3$ 个随机种子），并报告代表性数值 $M = 16$、128、256；$M \leq 256$ 因为超过 $M \approx 300$ 后，2vR 列的数量会变得难以实际应用。