

One-against-all

Daniel Hsu (COMS 4771)

Theorem. Let $\hat{\eta}_1, \dots, \hat{\eta}_K: \mathcal{X} \rightarrow [0, 1]$ be estimates of conditional probability functions $x \mapsto \mathbb{P}(Y = k \mid X = x)$ for $k = 1, \dots, K$, and let

$$\epsilon := \mathbb{E} \left[\max_{k=1, \dots, K} \left| \hat{\eta}_k(X) - \mathbb{P}(Y = k \mid X) \right| \right].$$

Let $\hat{f}: \mathcal{X} \rightarrow \{1, \dots, K\}$ be the one-against-all classifier based on $\hat{\eta}_1, \dots, \hat{\eta}_K$, i.e.,

$$\hat{f}(x) = \arg \max_{k=1, \dots, K} \hat{\eta}_k(x), \quad x \in \mathcal{X},$$

(with ties broken arbitrarily), and let $f^*: \mathcal{X} \rightarrow \{1, \dots, K\}$ be the Bayes optimal classifier. Then

$$\mathbb{P}(\hat{f}(X) \neq Y) - \mathbb{P}(f^*(X) \neq Y) \leq 2\epsilon.$$

Proof. Fix $x \in \mathcal{X}$, $y^* := f^*(x)$, and $\hat{y} := \hat{f}(x)$. Let $\eta_k(x) := \mathbb{P}(Y = k \mid X = x)$ for all $k = 1, \dots, K$. Then

$$\begin{aligned} \mathbb{P}(\hat{f}(X) \neq Y \mid X = x) &= \mathbb{P}(\hat{y} \neq y^* \mid X = x) \\ &= \mathbb{P}(\hat{\eta}_{y^*}(x) < \hat{\eta}_{\hat{y}}(x) \mid X = x) \\ &= \mathbb{P}(\hat{\eta}_{y^*}(x) - \eta_{y^*}(x) < \hat{\eta}_{\hat{y}}(x) - \eta_{\hat{y}}(x) \mid X = x) \\ &\leq \mathbb{P}(\max_{k=1, \dots, K} |\hat{\eta}_k(x) - \eta_k(x)| > 0 \mid X = x) \end{aligned}$$

Therefore, taking expectations with respect to X ,

$$\mathbb{P}(\hat{f}(X) \neq Y) - \mathbb{P}(f^*(X) \neq Y) \leq 2 \cdot \mathbb{E} \left[\max_{k=1, \dots, K} |\hat{\eta}_k(X) - \eta_k(X)| \right].$$

The bound on the excess risk is tight. To see this, suppose for a given $x \in \mathcal{X}$ (with $y^* = f^*(x)$ and $\hat{y} = \hat{f}(x)$), we have $\hat{\eta}_{y^*}(x) = \hat{\eta}_{\hat{y}}(x) - \delta$, but $\eta_{y^*}(x) = \hat{\eta}_{y^*}(x) + \epsilon$ and $\eta_{\hat{y}}(x) = \hat{\eta}_{\hat{y}}(x) - \epsilon$. Then

$$\begin{aligned} \eta_{y^*}(x) - \eta_{\hat{y}}(x) &= (\hat{\eta}_{y^*}(x) + \epsilon) - (\hat{\eta}_{\hat{y}}(x) - \epsilon) \\ &= 2\epsilon - \delta \end{aligned}$$

which tends to 2ϵ as $\delta \rightarrow 0$.