Supplemental material: How many faces can it recognize? Performance extrapolation for multi-class classification

Charles Zheng, Rakesh Achanta and Yuval Benjamini

May 19, 2016

1 Proofs

Theorem 3.1. *Let $\mathcal{Q}$ be a single-distribution classification function, and let $\mathbb{F}$, $\hat{\mathbb{F}}(F)$ be a distribution on $\mathcal{P}(\mathcal{Y})$. Further assume that $\hat{\mathbb{F}}$ and $\mathcal{Q}$ jointly satisfy the* tie-breaking *property:*

$$\Pr[\mathcal{Q}(\hat{F}, y) = \mathcal{Q}(\hat{F}', y)] = 0 \tag{1}$$

for all $y \in \mathcal{Y}$, where $\hat{F}, \hat{F}' \stackrel{iid}{\sim} \hat{\mathbb{F}}$. Let U be defined as the random variable $U = u(\hat{F}, Y)$ for $F \sim \mathbb{F}$, $Y \sim F$, and $\hat{F} \sim \hat{\mathbb{F}}(F)$ with $Y \perp \hat{F}$. Then

$$p_k = \boldsymbol{E}[U^{k-1}],$$

*where p_k is the expected accuracy as defined by (**??**).*

Proof. Write $q^{(i)}(y) = \mathcal{Q}(\hat{F}_i, y)$. By using conditioning and conditional independence, p_k can be written

$$\begin{aligned}
p_k &= \mathbf{E}\left[\frac{1}{k}\sum_{i=1}^{k}\Pr_{F_i}[q^{(i)}(Y) > \max_{j\neq i} q^{(j)}(Y)]\right] \\
&= \mathbf{E}\left[\Pr_{F_1}[q^{(1)}(Y) > \max_{j\neq 1} q^{(j)}(Y)]\right] \\
&= \mathbf{E}_{F_1}[\Pr[q^{(1)}(Y) > \max_{j\neq 1} q^{(j)}(Y)|\hat{F}_1, Y]] \\
&= \mathbf{E}_{F_1}[\Pr[\cap_{j>1} q^{(1)}(Y) > q^{(j)}(Y)|\hat{F}_1, Y]] \\
&= \mathbf{E}_{F_1}[\prod_{j>1}\Pr[q^{(1)}(Y) > q^{(j)}(Y)|\hat{F}_1, Y]] \\
&= \mathbf{E}_{F_1}[\Pr[q^{(1)}(Y) > q^{(2)}(Y)|\hat{F}_1, Y]^{k-1}] \\
&= \mathbf{E}_{F_1}[u(\hat{F}_1, Y)^{k-1}] = \mathbf{E}[U^{k-1}].
\end{aligned}$$

□