Modern Multivariate Statistical Techniques

Jinhong Du,15338039

April 30, 2018

Content

1. For any split of τ into τ_L and τ_R,
$$R^{re}(\tau) \geq R^{re}(\tau_L) + R^{re}(\tau_R)$$
with equality if $\arg\max_k p(k|\tau) = \arg\max_k p(k|\tau_L) = \arg\max_k p(k|\tau_R)$.

Proof.

Denote by τ_L and τ_R the left daughter-node and right daughter-node, respectively, emanating from a (parent) node τ.

$r(\tau)$ is the resubstitution estimate of the misclassification rate $R^{re}(\tau)$ of an observation in node τ.

$p(k|\tau)$ is an estimate of $\mathbb{P}\{X \in \prod_k |\tau\}$, the conditional probability that an observation X is in $\prod_k$ given that it falls into node τ.

$p(\tau)$ is the proportion of all observations that fall into node τ.

$\because$

$$\begin{aligned}
r(\tau) &= 1 - \max_k p(k|\tau) \\
r(\tau_L) &= 1 - \max_k p(k|\tau_L) \\
r(\tau_R) &= 1 - \max_k p(k|\tau_R) \\
R^{re}(\tau) &= r(\tau)p(\tau) \\
&= r(\tau)[p(\tau_L) + p(\tau_R)] \\
R^{re}(\tau_L) &= r(\tau_L)p(\tau_L) \\
R^{re}(\tau_R) &= r(\tau_R)p(\tau_R)
\end{aligned}$$

$\therefore$

$$\begin{aligned}
R^{re}(\tau) - R^{re}(\tau_L) - R^{re}(\tau_R) &= [r(\tau) - r(\tau_L)]p(\tau_L) + [r(\tau) - r(\tau_R)]p(\tau_R) \\
&= [\max_k p(k|\tau_L) - \max_k p(k|\tau)]p(\tau_L) + [\max_k p(k|\tau_R) - \max_k p(k|\tau)]p(\tau_R)
\end{aligned}$$

$\because$

$$\begin{aligned}
p(k|\tau) &= \frac{p(\tau_L)p(k|\tau_L) + p(\tau_R)p(k|\tau_R)}{p(\tau)} \\
\max_k p(k|\tau) &\leq \frac{p(\tau_L)}{p(\tau)} \max_k p(k|\tau_L) + \frac{p(\tau_R)}{p(\tau)} \max_k p(k|\tau_R) \qquad (1)
\end{aligned}$$

$\therefore$

$$\begin{aligned}
R^{re}(\tau) - R^{re}(\tau_L) - R^{re}(\tau_R) \geq & \frac{p(\tau_R)p(\tau_L)}{p(\tau)}[\max_k p(k|\tau_L) - \max_k p(k|\tau_R)] \\
& + \frac{p(\tau_R)p(\tau_L)}{p(\tau)}[\max_k p(k|\tau_R) - \max_k p(k|\tau_L)]
\end{aligned}$$

$$= 0$$

The equality holds if $\arg\max_k p(k|\tau) = \arg\max_k p(k|\tau_L) = \arg\max_k p(k|\tau_R)$ from (1). $\square$

2. Show that the entropy function of $p(1|r), \cdots, p(K|r)$,

$$\begin{aligned} i(r) &= \phi(p(1|r), \cdots, p(K|r)) \\ &= -\sum_{k=1}^{K} p(k|r) \log p(k|r) \end{aligned}$$

is maximized at $\left(\frac{1}{K}, \cdots, \frac{1}{K}\right)$.

Proof.

Assume that $x \log x|_{x=0} = 0$.

$\because$

$$\begin{aligned} f(x) &= x \log x, \qquad x \in [0,1] \\ f'(x) &= \log x + 1, \qquad x \in [0,1] \\ f''(x) &= \frac{1}{x} \geq 0, \qquad x \in [0,1] \end{aligned}$$

$\therefore$ $f(x)$ is a convex function on $[0,1]$

$\because$

$$\sum_{k=1}^{K} p(k|r) = 1$$

$\therefore$ from Jensen Inequlity,

$$\begin{aligned} f\left(\frac{1}{K}\sum_{k=1}^{K} p(k|r)\right) &\leq \frac{1}{K}\sum_{k=1}^{K} f(p(k|r)) \\ f\left(\frac{1}{K}\right) &\leq -\frac{1}{K} i(r) \end{aligned}$$

$\therefore$

$$\begin{aligned} i(r) &\leq -f\left(\frac{1}{K}\right) \\ &= -\frac{1}{K}\log\frac{1}{K} \\ &= \frac{1}{K}\log K \end{aligned}$$

The equality holds when $p(1|r) = \cdots = p(K|r) = \frac{1}{K}$.

$\square$