Appendix – Proofs of Theorems 1-3

Proof of Theorem 1

Let $\hat{C} = I(T(X) \geq u)$ and consider the PPV of this classifier,

$$PPV(u) = \frac{\pi_1 P(T(X) \geq u \mid C = 1)}{\pi_1 P(T(X) \geq u \mid C = 1) + \pi_0 P(T(X) \geq u \mid C = 0)}.$$

We will first show $PPV(u)$ is an increasing function of $u$. By equivalently showing that $r(u) = \frac{P(T(X) \geq u \mid C = 1)}{P(T(X) \geq u \mid C = 0)}$ is an increasing function of $u$. With $T(X) = \log\left(\frac{h_1(X)}{h_0(X)}\right)$,

$$r(u) = \frac{\int_{T(X) \geq u} h_1(x) \, dx}{\int_{T(X) > u} h_0(x) \, dx} = \frac{\int_{T(X) \geq u} \left(\frac{h_1(x)}{h_0(x)}\right) h_0(x) \, dx}{\int_{T(X) > u} h_0(x) \, dx} = E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, T(X) \geq u\right\}.$$

For any $u_1 < u_2$, it follows from the law of total probability that

$$r(u_1) = E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, T(X) \geq u_1\right\}$$

$$= P\{T(X) \geq u_2 \mid C = 0, T(X) \geq u_1\} \times E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, T(X) \geq u_2\right\}$$

$$+ P\{u_1 \leq T(X) < u_2 \mid C = 0, T(X) \geq u_1\} \times E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, u_1 \leq T(X) < u_2\right\}$$

$$= \omega r(u_2) + (1 - \omega) r^*$$

where $\omega = P\{T(X) \geq u_2 \mid C = 0, T(X) \geq u_1\}$ and $r^* = E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, u_1 \leq T(X) < u_2\right\}$. By definition,

$$r(u_2) = E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, T(X) \geq u_2\right\} \geq e^{\mu_2} \geq E\left\{\frac{h_1(x)}{h_0(x)} \mid C = 0, u_1 < T(X) < u_2\right\} = r^*.$$
Because \( r(u_i) \) is a weighted average of \( r(u_2) \) and \( r' \), this implies \( r(u_i) \leq r(u_2) \) and this shows that \( r(u) \) and hence \( PPV(u) \) are increasing in \( u \). Similarly it can be shown that \( NPV(u) \) is decreasing in \( u \) and it therefore follows that the PROC is monotone.

**Proof of Theorem 2**

If \( T(X) \) is the log-likelihood ratio then Theorem 1 implies the PROC curve is increasing, which implies\(^{16} \) that \( h_o(u) \leq h_p(u) \) and \( \bar{h}_p(u) \leq \bar{h}_o(u) \) hold. The theorem then follows from necessary and sufficient conditions given in section 4.

**Proof of Theorem 3**

Define \( R_1 \) to be the set of all \((x_1, x_2)\) at the second stage for which the classifier predicts class 1, and let \( R_1^c \) denote the complement of \( R_1 \). Then,

\[
OER_2(u) = \pi_0 FPR_2(u) + \pi_1 FNPR_2(u) \\
= \pi_0 P(T_1(X_i) \geq u_i | C = 0) + \pi_0 P(u_0 < T_1(X_i) < u_i, (X_1, X_2) \in R_1 | C = 0) \\
+ \pi_1 P(T_1(X_i) < u_0 | C = 1) + \pi_1 P(u_0 < T_1(X_i) < u_i, (X_1, X_2) \in R_1^c | C = 1) \\
= \pi_0 P(T_1(X_i) \geq u_i | C = 0) + \pi_1 P(T_1(X_i) < u_0 | C = 1) + \pi_1 P(u_0 < T_1(X_i) < u_i | C = 1) \\
+ \left\lfloor \int_{R_1} I\{u_0 < T_1(x_i) < u_i\} \{\pi_0 h_0(x_1, x_2) - \pi_1 h_1(x_1, x_2)\} dx_1 dx_2 \right\rfloor,
\]

It is clear that to minimize \( OER_2(u) \) we should take \( R_1 \) to include all points \((x_1, x_2)\) for which \( h_1(x_1, x_2) / h_0(x_1, x_2) > \pi_0 / \pi_1 \), which is equivalent to the claim in the theorem.