STAT 672: Homework 2

Tom Wallace

February 20, 2018

Problem 1

For a supervised learning problem, risk is defined as the expected value of the loss function:

$$R(f) = \mathbf{E}_{X,Y\sim P}[L(Y, f(x))]$$

Bayes risk is the risk present when using the Bayes classifier:

$$f^*(x) = \arg\max_{Y} P(Y = y|X = x)$$

Typically, this classifier is not practical because we do not know the conditional distribution of Y given X, but in this problem it is given. Our resultant Bayes classifier is:

$$f^*(x) = \begin{cases} 0 & x \in [0.2, 0.8] \\ 1 & x \in \{(0, 0.2) \cup (0.8, 1)\} \end{cases}$$

Suppose that we use a typical 0—1 loss function.

$$L(y, f(x)) = \begin{cases} 0 & y = \text{sign}(f(x)) \\ 1 & \text{otherwise} \end{cases}$$

Thus, risk in our problem is equal to:

$$P(\text{sign}(Y) \neq \text{sign}(f(x)))$$

Using the law of total probability, this is equal to:

$$P(Y = 0|X \in \{(0, 0.2) \cup (0.8, 1)\})P(X \in \{(0, 0.2) \cup (0.8, 1)\}) + P(Y = 1|X \in [0.2, 0.8])P(X \in [0.2, 0.8])$$

$$(0.2 \times 0.1) + (0.2 \times 0.1) + (0.2 \times 0.6) = 0.16$$