5.2

Intro

In this exercise we will perform optimal Bayesian decision theory on the binary classificaiton problem with arbitrary misclassification losses. We will assume that $y \in \{0, 1\}$. Intuitively, we expect that when the loss λ_{10} (predicting one when the class is zero) is small, the threshold for choosing 1 will also be small. A similar argument can be made for the 0 class. Both this problem and problem 5.4 show very useful applications of Bayesian decision theory, so I think is worth for all to take a look.

Solution

a)

In a Bayesian decision problem, our goal is to find the Bayes rule that **minimizes the posterior expected loss** $\rho(a|x)$. In that case, let's write it down the expression for our expected loss:

$$\begin{aligned}&\rho(a|x) = E_{p(y|x)}[L(y,a)] = p_0 L(0,a) + p_1 L(1,a) = \\ &p_0 L(0,\hat{y}) + p_1 L(1,\hat{y}) = p_0 \lambda_{10} \hat{y} + p_1 \lambda_{01}(1-\hat{y})\end{aligned} \tag{1}$$

From Equation 1, we conclude that in order to minimize the posterior loss, we have to choose action $a = \hat{y}$ such that:

$$\begin{aligned}&\min \rho(a|x) = \min p_0 \lambda_{10} \hat{y} + p_1 \lambda_{01}(1-\hat{y}) = \\ &p_1 \lambda_{01} \iff p_1 \lambda_{01} < p_0 \lambda_{10} = (1-p_1)\lambda_{10} \iff p_1 < \frac{\lambda_{10}}{\lambda_{01}+\lambda_{10}}\end{aligned} \tag{2}$$

Therefore, the threshold for action selection is $\theta = \frac{\lambda_{10}}{\lambda_{01}+\lambda_{10}}$. Expressing this statement in the form of a action function, we arrive at:

$$a = \delta(x) = \begin{cases} 1 \text{ if } p_1 \geq \theta \\ 0 \text{ elsewhere} \end{cases} \tag{3}$$

b)

Our threshold is 0.1. Thus, $\theta = \frac{\lambda_{10}}{\lambda_{01}+\lambda_{10}} = 0.1$. Therefore, we have the following relationship among the misclassification losses:

$$0.1 = \frac{\lambda_{10}}{\lambda_{01}+\lambda_{10}} \iff \lambda_{01} = 9\lambda_{10} \tag{4}$$

Equation 4 gives the only relation we need to build the loss matrix. The reason we only need Equation 4 and not fixed values is that it does not matter the absolute value of either one of the misclassification rates when they are the only variables in the problem. The only thing that matters is their value with respect with each other. Therefore, choosing $\lambda_{10} = 1$, we have:

Table 1: Loss matrix

predicted label ŷ	true label y	
	0	1
0	0	9
1	1	0

Conclusion

In this exercise, we found the optimal action policy $\delta(x)$ for a binary classification problem with arbitrary misclassification losses. As suspected in the introduction, we discovered that the threshold $\theta = \frac{\lambda_{10}}{\lambda_{01}+\lambda_{10}}$ for selection p_1 is small when the misclassification λ_{10} is also small (compared to λ_{01}). In item b, we analysed a concrete example, where we built a loss matrix from a given threshold. We observed that the absolute values of the misclassification losses didn't matter. What really matters is their relative size with respect with each other.