5.3

Intro

In this exercise we will perform optimal bayesian decision theory in a classification problem with the possibility of rejection. Quoting Tyrion Lannister: "...Sometimes nothing is the hardest thing to do... ". Jokes aside, doing nothing always have a cost, as we may well know. But the cost can be much smaller than taking a wrong decision. In item a we will derive the the optimal choice, as a function of the costs λ_r and λ_s. Our intuition says that if the cost of rejection λ_r is small compared to λ_s, we will have a strong bias towards rejection. On the other hand, if λ_r is big compared to λ_s, we will only choose rejection in extreme situations. An interesting discussion about this type of behaviour will be made in item b.

Solution

a)

Let's start by writing down the posterior expected loss:

$$\rho(a|x) = E_{p(y|x)}[L(y,a)] = \sum_{k=1}^{C} p(y=k|x)L(y=k,a) \tag{1}$$

Let's us make a choice α_d. If $\alpha_d \in \{\alpha_1, \alpha_2, ..., \alpha_C\}$, then Equation 1 becomes:

$$\begin{aligned} \rho(a|x) = \sum_{k=1}^{C} p(y=k|x)L(y=k,\alpha_d) = \\ \sum_{k\neq d} p(y=k|x)\lambda_s = (1-p(y=d|x))\lambda_s \end{aligned} \tag{2}$$

Our other choice is to choose to reject all classes. In this case, Equation 1 becomes:

$$\begin{aligned} \rho(a|x) = \sum_{k=1}^{C} p(y=k|x)L(y=k,\alpha_d) = \\ \sum_{k=1}^{C} p(y=k|x)\lambda_r = \lambda_r \end{aligned} \tag{3}$$

Now, the question is what choice minimize the posterior expected loss. Assuming the right choice among all classes is $Y=j$ we have:

$$\begin{aligned} (1-p(y=j|x))\lambda_s \leq (1-p(y=d|x))\lambda_s \\ p(y=j|x) \geq p(y=d|x) \;\; d=1,2,...,C \end{aligned} \tag{4}$$

This guarantees that $Y=j$ is the best choice among the classes. Now, we need to know if it is best than the reject option.

$$\begin{aligned} (1-p(y=j|x))\lambda_s \leq \lambda_r \\ p(y=j|x) \geq 1-\frac{\lambda_r}{\lambda_s} \end{aligned} \tag{5}$$

b)

Let's begin by discussing the extreme cases. If $\frac{\lambda_r}{\lambda_s} = 0$, then there is no risk in rejection. Therefore, the only case we will not reject is when we have 100% sure that class $Y = j$ is the right one. Using Equation 5, this means: $p(y = j|x) \geq 1 - \frac{\lambda_r}{\lambda_s} = 1$. If $\frac{\lambda_r}{\lambda_s} = 1$, this means that rejecting all options is as bad as misclassification. Therefore, it is better to choose something and have at least a chance of getting the right class. Using Equation 5, this means: $p(y = j|x) \geq 1 - \frac{\lambda_r}{\lambda_s} = 0$. In the middle of the extremes cases, as the relative cost $\frac{\lambda_r}{\lambda_s}$ increases, it becomes less and less interesting to choose rejection. This makes sense, because the cost of rejection is becoming as bad as the cost of misclassification. Another important point is that $\lambda_r \leq \lambda_s$ in other for us to even consider the option of rejection.

Conclusion

In this exercise we performed bayesian optimal decision theory in a classification problem with reject option. In item a, we discovered the optimal action policy, which tell us when to choose the best class among the C options and when to reject all classes. In item b, we discussed how our action policy changes if function of $\frac{\lambda_r}{\lambda_s}$. We discovered that our policy change accordingly with our intuitions stated at the introduction.