Machine Learning: Excercise 1

Sergio de las Heras Perez, Mario García Marquez,
Daniel Quinteiro Donaghy, Jorge Rey Melián
436857 , 449118, 450129, 450067

7 November 2022

2 Question 2: Warming Up!

2.a) What is the probability of selecting an apple?

Since before picking an apple we need to pick a box, we need to know the probability of picking an apple from each box. For this we are using the formula

$$p(apple|Box) = \frac{N_{apples}}{N_{fruit}} \tag{1}$$

Where N_{apples} and N_{fruit} is specific for each box. Since $\{r, g, b\}$, which denotes the events of picking the red, green or blue box are a partition of this event space, we can calculate the probability of picking an apple as follows:

$$p(apple) = \sum_{j}^{\{r,g,b\}} p(apple|j)p(j) = 0.2\frac{3}{10} + 0.2\frac{1}{2} + 0.6\frac{3}{10} = 0.34 \tag{2}$$

2.b) If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Here we can use the Bayes' Theorem which states that

$$p(x|\alpha) = \frac{p(\alpha|x)p(x)}{p(\alpha)} \tag{3}$$

We want to calculate $p(g|orange)$. So we can calculate the rest of the necessary terms for applying the formula.

$$p(orange|g) = \frac{3}{10} \tag{4}$$

$$p(orange) = \sum_{j}^{\{r,g,b\}} p(orange|j)p(j) = 0.2\frac{4}{10} + 0.2\frac{1}{2} + 0.6\frac{3}{10} = 0.36 \tag{5}$$

So the solution to our problem is:

$$p(g|orange) = \frac{p(orange|green)p(green)}{p(orange)} = 0.5 \tag{6}$$

3 Question 3: Minimizing the Expected Loss

3.a) Show that the minimum risk is obtained when

$$\text{classify}(x) :\rightarrow \begin{cases} C_j, & \text{iff} \quad \forall k : p(C_j|x) \geq p(C_k|x) \wedge p(C_j|x) \geq 1 - \frac{l_r}{l_s} \\ C_{rej}, & \text{otherwise} \end{cases} \tag{7}$$

We know that

$$L_{kj} = \begin{cases} 0 & \text{if } k = j \\ l_r & \text{if } j = N+1 \\ l_s & \text{otherwise} \end{cases} \tag{8}$$

Let's suppose that we have to classify a variable x to a class $C \in \{C_1, C_2, ..., C_N, C_{rej}\}$. In this case, we have that the risk(= expected loss) is

$$R(\alpha_i|x) = \sum_{j=1}^{N+1} L_{ji} p(C_j|x) \tag{9}$$

where α_i represents the decision of picking the class C_i. So in order to minimize the loss, we want to pick the decision with the minimum risk. Let's take a look to the loss matrix(in this case, with 3 classes for illustration)

$$\begin{pmatrix} 0 & l_s & l_s & l_r \\ l_s & 0 & l_s & l_r \\ l_s & l_s & 0 & l_r \end{pmatrix} \tag{10}$$

So, for three classes, the risk of picking C_1 would be

$$R(\alpha_1|x) = l_s p(C_2|x) + l_s p(C_3|x) + l_r \tag{11}$$

and more generally would be

$$R(\alpha_i|x) = l_s(p(C_1|x) + p(C_2|x) + ... + p(C_{i-1}|x) + p(C_{i+1}|x) + ... + p(C_N|x)) + l_r \tag{12}$$

so we will pick C_1 if $R(\alpha_1|x) < R(\alpha_i|x)$ for $i \neq 1$

$$R(\alpha_1|x) < R(\alpha_i|x) \iff p(C_1|x) < p(C_i|x) \tag{13}$$

(in order to verify this you just need to substitute $R(\alpha|x)$ for this expression). Now we need to check when is better to reject x.

$$R(\alpha_{rej}|x) = l_s(\sum_i^N p(C_i|x)) \tag{14}$$

So we got

$$R(\alpha_1|x) < R(\alpha_{rej}|x) \iff -l_s p(C_1|x) + l_r < l_s \iff p(C_1|x) > \frac{l_s - l_r}{l_s} \tag{15}$$

So unless this condition is meet, the optimal option is to reject x.

3.b) What happens if $l_r = 0$

If $l_r = 0$ then the threshold for rejecting will be higher: $1 - \frac{1}{ls}$ so we will be rejecting way more values than usual.

3.c) What happens if $l_r > l_s$

If $l_r > l_s$ we got that $1 - \frac{l_r}{l_s} < 0$ and then will never reject any value.

4 Question 4: Maximum Likelihood

Given the function

$$p(x|\theta) = \theta^2 x \exp(-\theta x) g(x) \tag{16}$$

Our goal is now to compute an estimation of the parameter $\tilde{\theta}$ from the observations $x_1, ..., x_n$ using the maximum likelihood method.

Assuming that all datapoints are independent from each other, we can write the likelihood of θ as

$$L(\theta) = p(X|\theta) = \prod_{n=1}^{N} p(x_n|\theta) \tag{17}$$

Or simplify this expression by taking the logarithm and computing the log-likelihood

$$E(\theta) = -\ln L(\theta) = -\sum_{n=1}^{N} \ln p(x_n|\theta) \tag{18}$$

Using our initial probability function, deriving and setting to zero we get:

$$E(\theta) = -\sum_{n=1}^{N} \ln(\theta^2 x_n \exp(-\theta x) g(x_n)) \tag{19}$$

$$\frac{\partial}{\partial\theta} E(\theta) = -\frac{\partial}{\partial\theta} \sum_{n=1}^{N} \ln(\theta^2 x_n \exp(-\theta x_n) g(x_n)) \stackrel{!}{=} 0 \tag{20}$$

The derivation can be introduced inside the sum and expanding the logarithm:

$$-\sum_{n=1}^{N} \frac{\partial}{\partial\theta} \ln(\theta^2 x_n \exp(-\theta x_n) g(x_n)) \stackrel{!}{=} 0 \tag{21}$$

$$-\sum_{n=1}^{N} \frac{\partial}{\partial\theta} (2\ln(\theta) + \ln(x_n) - \theta x_n - \ln(g(x_n))) \stackrel{!}{=} 0 \tag{22}$$

$$-\sum_{n=1}^{N} 2\frac{1}{\theta} - x_n - \stackrel{!}{=} 0 \tag{23}$$

$$-2N\frac{1}{\theta} \sum_{n=1}^{N} x_n \stackrel{!}{=} 0 \tag{24}$$

$$\tag{25}$$

And finally we get the estimate of $\tilde{\theta}$ as:

$$\tilde{\theta} = -2N \sum_{n=1}^{N} x_n \tag{26}$$