Exercise 1 (5 points): Discrete Naive Bayes Classifier [Pen and Paper]

In this exercise, we want to get a basic idea of the naive Bayes classifier by analysing a small example. Suppose we want to classify fruits based on the criteria length, sweetness and the colour of the fruit and we already spent days by categorizing 1900 fruits. The results are summarized in the following table.

	Length			Sweetness		Colour			
Class	Short	Medium	Long	Sweet	Not Sweet	Red	Yellow	Green	Total
Banana	0	100	500	500	100	0	600	0	600
Papaya	50	200	50	250	50	0	150	150	300
Apple	900	100	0	800	200	600	100	300	1000
Total	950	400	550	1550	350	600	850	450	1900

To keep our classifier simple, we deal only with three fruits (banana, papaya and apples). The general scenario is as follows: we get a new fruit where we do not know its class. However, we can measure its length, take a taste and name its colour. Based on these features (the criteria) we then want to know which type of fruit we have. The features are assumed to be independent.

1. Regarding the features and their level of measurement: what scale do they have (i.e. nominal-, ordinal-, interval- or ratio-scaling)?

2. What is the prior probability of a banana, i.e. $P(Banana)$?

3. What is the prior probability of a long fruit, i.e. $P(Long)$?

4. What is the probability that you have a banana in your hand, when you already know that the fruit is long, i.e. $P(Banana|Long)$?

5. What is the probability that the fruit is long, when you already know that it is a banana (the likelihood of the evidence), i.e. $P(Long|Banana)$?

6. The naive Bayes classifier iterates over every possible class and calculates the conditional probability that the class fits the known features. The class with the maximum corresponding probability is selected

$$\text{argmax}_{\omega \in \{Banana,\ Papaya,\ Apple\}} \left(P(\omega|x) \right)$$

with

$$P(\omega|x) = \frac{P(x_1|\omega)\cdots P(x_n|\omega) \cdot P(\omega)}{p(x)}.$$

For a fruit which has medium length, tastes sweet and looks green to you (i.e. $x = (\textit{Medium}, \textit{Sweet}, \textit{Green})^T$), which class does it most likely belong to? Hint: you do not need to calculate $p(x)$.

Exercise 2 (5 points): Misclassification Costs

With the Bayes decision rule, we distinguish between classes solely based on probability distributions. This rule is fixed in the sense that we have no manual control over the decision process. But not every class is of the same importance and sometimes we wish to set an individual focus. This can be achieved with penalty terms and is the topic of this exercise.

We consider a two-class problem and use the normal distributions from the example in the lecture script (slide 27) as likelihood functions

$$p(x|\omega_1) = \frac{1}{\sqrt{\pi}} \exp(-x^2)$$

$$p(x|\omega_2) = \frac{1}{\sqrt{\pi}} \exp(-(x-1)^2).$$

We do not set concrete values for the a-priori probabilities, yet. But as we only consider two classes, we can use the variable $p \in [0; 1]$ to describe both

$$P(\omega_1) = p \quad \text{and} \quad P(\omega_2) = 1 - p.$$

Now, we have all ingredients together to define the loss of each class

$$l_1(x) = \lambda_{21} p(x \mid \omega_2) P(\omega_2)$$
$$l_2(x) = \lambda_{12} p(x \mid \omega_1) P(\omega_1)$$

where correct classifications are not penalized, i.e. $\lambda_{ii} = 0$. The term $\lambda_{ij} \in \mathbb{R}^+$ weights the probabilities and penalizes if a pattern which belongs to the class ω_i is misclassified to the class ω_j.

To assign patterns to classes we use a similar decision rule as in the Bayes case. The only difference is that we select the class which yields the lowest loss instead of the highest probability

$$\omega^* = \text{argmin}_i (l_i(x)). \tag{1}$$

In the case of our two-class problem, we assign a new pattern x to ω_1 if $l_1(x) < l_2(x)$ and to ω_2 if $l_2(x) < l_1(x)$.

1. [Pen and Paper] Explain why it is reasonable that the loss of the second class $l_2(x)$ uses the probabilities of the first class.

2. [Pen and Paper] Suppose we have $l_1(x) = 2$ and $l_2(x) = \lambda_{12} \cdot 3$. Define the value(s) of λ_{12} which correspond to the following cases:

   a) The pattern x is assigned to the first class ω_1.

   b) The pattern x is assigned to the second class ω_2.

   c) The decision is ambiguous.

3. [Pen and Paper] The advantage of a two-class problem is that we can reduce the decision of Equation 1 to a simple threshold $\hat{x}_0$ separating the regions of the two classes. Find an appropriate ansatz and show that the threshold is given by

$$\hat{x}_0(\lambda_{12}, \lambda_{21}, p) = \frac{1}{2}\left(\ln\left(\frac{\lambda_{12}}{\lambda_{21}} \cdot \frac{p}{1-p}\right) + 1\right).$$

4. [Python] Write a script which plots the likelihood functions for the parameters $\lambda_{12} = \frac{1}{2}, \lambda_{21} = 1$ and $p = \frac{1}{2}$ and mark the threshold $\hat{x}_0$ as a vertical line in the graph. Plot also the losses l_1 and l_2 and mark the threshold value $\hat{x}_0$ here as well. An example plot is shown in Figure 1.

5. [Python] Change the parameters to $\lambda_{12} = \frac{7}{2}, \lambda_{21} = 1$ and $p = \frac{1}{3}$ and plot the result.

6. [Python] Change the parameters to $\lambda_{12} = 7, \lambda_{21} = 2$ and $p = \frac{1}{3}$ and plot the result. Does the threshold change compared to the previous settings? If not, why?

7. [Python] Change the parameters to $\lambda_{12} = 1$ and $\lambda_{21} = 1$. For which value(s) of p will the threshold line in the likelihood functions be

   a) left of the intersection $p(x|\omega_1) = p(x|\omega_2)$,

   b) at the intersection $p(x|\omega_1) = p(x|\omega_2)$,

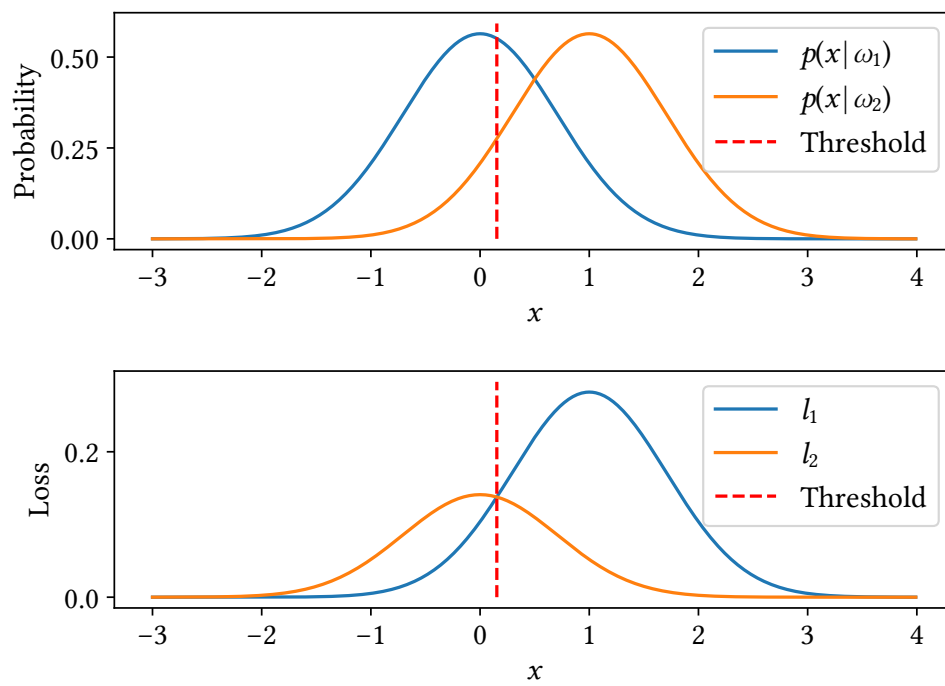   c) right of the intersection $p(x|\omega_1) = p(x|\omega_2)$?

Figure 1: Likelihood functions and losses corresponding to the two classes of Exercise 2.