Machine Learning Exercises 8

Exercise 1. Revisit the 10-nearest neighbour classifier that you made in Exercises set 7 based on the dataset Ex1-training.csv with two features (x_1 and x_2) and three classes (Black = 1, Red = 2, Blue = 3).

In fact, the provided training and test data were balanced – there were equally many instances in each of the three classes – and this distribution of class instances does not reflect that of future datasets to which we wish to apply the classifier. Informed by other studies we know that the prior probabilities of classes would be

$$\mathrm{P}(Y = y) = \begin{cases} 0.0001, & \text{if } y = \text{Black} \\ 0.02, & \text{if } y = \text{Red} \\ 0.979, & \text{if } y = \text{Blue} \end{cases}$$

Modify your 10-nearest-neighbour classifier to reflect this prior class distribution and visualise the resulting decision regions. Describe how they have changed compared to the visualisation you made in Exercise set 7.

Exercise 2. Assuming that you are given an expression for $p(x, y)$, explain how you can achieve the marginal distributions, $p(y)$ and $p(x)$, as well as the conditional distributions $p\,(y|x)$ and $p\,(x|y)$.

Exercise 3. Imagine that you wish to detect whether a person has cancer ($Y = 1$) from some continuous measurement, e.g. a blood marker.

Assume that the class conditionals are normal distributions as

$$X \mid Y = 1 \sim \mathcal{N}(3, 2) \text{ and } X \mid Y = 2 \sim \mathcal{N}(1, 4)$$

and that class priors are $\mathrm{P}(Y = 1) = 0.2$ and $\mathrm{P}(Y = 2) = 0.8$.

1. Explain how to find Bayes classifier and illustrate on a plot both the decision regions and the two functions
$$g_k(x) = p\,(x|Y = k)\,\mathrm{P}(Y = k), k = 1, 2.$$
2. Normalise the plot above to show for each value of x the posterior class probabilities $\mathrm{P}\,(Y = k|X = x)$.
3. What is the Bayes error rate, i.e. the misclassification error for the Bayes classifier?
4. How does the classifier change, if instead $\mathrm{P}(Y = 1) = 0.5$?

Exercise 4. Consider again the model of Exercise 3. Now we change the loss function from misclassification error to a loss function specified by the following matrix

$$L = \begin{bmatrix} 0 & 1000 \\ 1 & 0 \end{bmatrix}$$

1. Explain why you may wish to use a more complex loss function than 0-1 loss.
2. Explain how to find a classifier that minimises the expected loss under this new loss function. *(Hint: use the posterior class probabilities from Exercise 3)*

Exercise 5. Now, in practice you would not know the Gaussian parameters, nor the class priors, so you need to add an inference step.

1. Simulate 1000 observations from the model in Exercise 3 and train the (plug-in) Bayes classifier on these.
2. Simulate a test data set of 200 observations and find the confusion matrix for your classifier.