

## Machine Learning Exercises 9

LDA and QDA are motivated by minimising the misclassification error (the expected loss under 0-1 loss) in a generative model where the class-conditionals are Gaussian (LDA: same variance across classes, QDA: different variance).

For a given value of the feature,  $x$ , the classifier works by computing  $K$  discriminant functions  $(\delta_1(x), \dots, \delta_K(x))$  and classifying to the class that has the highest value.

The expression for each  $\delta_k(x)$  is the quadratic expressions we derived for LDA and QDA jointly in lectures:

$$\delta_k(x) = -\frac{1}{2\sigma_k^2}x^2 + \frac{\mu_k}{\sigma_k^2}x - \frac{\mu_k^2}{2\sigma_k^2} - \frac{1}{2}\log(2\pi\sigma_k^2) + \log\pi_k. \quad (1)$$

In the case of LDA this reduces quite a bit since  $\sigma_k^2 = \sigma^2$  is the same for all classes.

To evaluate the discriminant functions, we need the Gaussian parameters and also the class probabilities. When these parameters are unknown, we estimate them from training data.

**Exercise 1.** In (1), show how this expression reduces in the case of LDA where all classes have the same variance. Explain how some terms can be eliminated and therefore that we can use *linear* rather than quadratic discriminants.

**Exercise 2.** Revisit the LDA example from Lectures 8, where there were three classes and a single continuous feature. Within each class the feature  $x$  has a univariate Gaussian distribution as

$$\begin{aligned} p(x|Y = \text{black}) &= \mathcal{N}(2, 1) \\ p(x|Y = \text{red}) &= \mathcal{N}(4, 1) \\ p(x|Y = \text{blue}) &= \mathcal{N}(7, 1), \end{aligned}$$

and the class probabilities are  $(\pi_{\text{black}}, \pi_{\text{red}}, \pi_{\text{blue}}) = (0.6, 0.1, 0.3)$ .

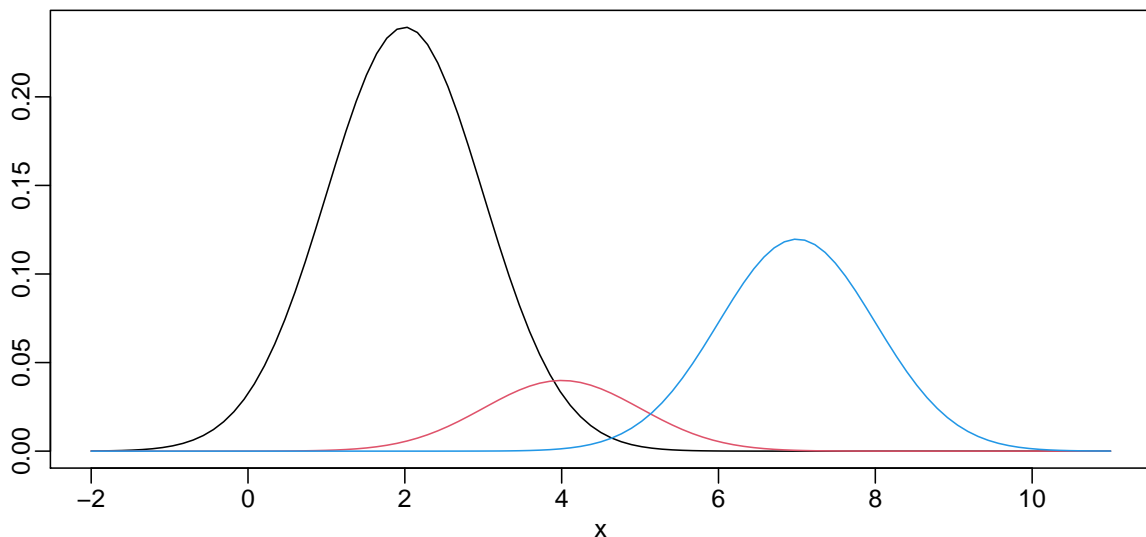


Figure 1: Visualisation of the joint distribution of feature and class.

a) Sketch on Figure 1 the decision regions for Bayes classifier.

- b) How many different misclassification errors can be made for this classification problem?
- c) For each possible error type, sketch on Figure 1 the area that represents the probability of making this error.
- d) Derive the exact decision boundaries for Bayes classifier. *Hint: use the discriminant functions for LDA, as obtained in Exercise 1 from reducing (1).*
- e) Compute the probabilities from c). *Hint: the Gaussian cumulative distribution function implements the integral under a Gaussian density function.*
- f) Compute the exact expected loss for the Bayes classifier.
- g) Simulate 1000 observations from the model and split into a training dataset with 600 observations and a test set with 400 observations.
- h) Train the classifier on the simulated data. This involves both finding the parameter estimates and computing the linear discriminant functions.
- i) Use the trained classifier to classify the test data and compute the confusion matrix (for each possible combination of true class and predicted class, how big a proportion of the 400 test data points end up in this combination?) and also the total misclassification error. Compare to the exact errors you got in e) and the Bayes error (misclassification error for the Bayes classifier) from f).
- j) Imagine we knew the relevant class probabilities (0.6, 0.1, 0.3) and could specify them directly in the LDA discriminant function rather than having to estimating them from the training data. Argue why we could hope to get a better classifier by training on a balanced dataset, rather than a dataset that is representative of the class probabilities.

**Exercise 3.** Now consider the QDA setup from lectures where the features were Gaussian but with each class having its own set of parameters (mean and variance):

$$\begin{aligned}
 p(x|Y = \text{black}) &= \mathcal{N}(2, 0.25) \\
 p(x|Y = \text{red}) &= \mathcal{N}(4, 1) \\
 p(x|Y = \text{blue}) &= \mathcal{N}(7, 0.81),
 \end{aligned}$$

- a) Explain why you would not expect LDA to perform well.
- b) Simulate some data from the model: A training dataset with 600 observations and a test set with 400 observations. Train both an LDA and a QDA classifier and compare their test error.