Tutorial 5 LDA, QDA, Naive Bayes and KNN

Introduction to Topic 4

Lecture

Suppose an anesthetist needs to determine whether an anesthetic is safe for a patient who is having a heart operation. The anesthesiologist may know certain things about this patient such as their age, gender, race, blood pressure and weight. Based on these kinds of data, the anesthesiologist would like to identify if the patient is safe or unsafe for anesthetic. This is a classification problem, we can employ one of many **Classification methods** for this. Last week, we started on logistic regression, this week we explore the use of Linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and K nearest approach (KNN)

Both LDA and QDA are based on the Gaussian mixture model framework, which assumes that the data is a mixture of K normal distributions, where K is the number of components. The probability of x_i belong to component k is,

$$\Pr(y = k|x_i) = \frac{\pi_k f_k(x_i)}{\sum_{l=1}^{K} \pi_l f_l(x_i)}$$

where π_k is the "weight" of the component and $f_k()$ is a Gaussian distribution with mean of μ_k and variance σ_k.

The key difference between LDA and QDA is in the variance term. In LDA, it assumes variance is the <u>same</u> for all classes (or components) while QDA allows covariance matrix to vary between classes (i.e. k). This makes the boundaries (discrimiinant function) between classes more flexible (in quadratic form).

The discriminant function for LDA is

$$\hat{\delta}_k(x) = x\frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k)$$

and QDA

$$\delta_k(x) = -\frac{1}{2}x^T\mathbf{\Sigma}_k^{-1}x + x^T\mathbf{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\mu_k^T\mathbf{\Sigma}_k^{-1}\mu_k + \log\pi_k$$

Here is just a snapshot of what we covered in the lecture. Now, it's your turn to work through the questions below in Group Two.

Independendent Learning

In a group of two, discuss the following topics

- Explain how LDA, QDA, Naive Bayes and KNN work
- What are the Key assumption behind LDA?
- What are the key difference between LDA and QDA?

- Why does QDA use class-specific covariance matrices?
- When would LDA be a better choice than QDA?
- What challenges might arise when using QDA with high-dimensional data?
- What does k represent in KNN?
- How does the choice of k affect the performance of the KNN algorithm?
- How would you choose an optimal value for k in KNN?
- Can KNN be used for regression tasks, and if so, how?

Not sure? Read these chapters

Author	Title	Relevant chapters
James et al	An Introduction to Statistical Learning with Applications in R	ch 4.4 ch 2.2.3

Tutorial

Independendent Learning

Labs 4.7 Lab: Classification Methods
4.7.1 The Stock Market Data
4.7.3 Linear Discriminant Analysis
4.7.4 Quadratic Discriminant Analysis
4.7.5 Naive Bayes
4.7.6 K-Nearest Neighbors

Exercises

- Question 2, Chapter 4.8
- Question 5, Chapter 4.8
- Question 7, Chapter 4.8
- Question 13, Chapter 4.8
- Question 13, Chapter 4.8