ch04_notes

August 10, 2019

Table of Contents
4 Logistic Regression
4.1 An Overview of Classification
4.2 Why Not Linear Regression?
4.3 Logistic Regression
4.3.1 The Logistic Model
4.3.2 Estimating the Regression Coefficients
4.3.3 Making Predictions
4.3.4 Multiple Logistic Regression
4.3.5 Logistic Regression for more than two response classes
4.4 Linear Discriminant Analysis
4.4.1 Bayes Theorem for Classification
4.4.2 Linear Discriminant Analysis for p=1
4.4.3 Linear Discriminant Analysis for p > 1
4.4.4 Quadratic Discriminant Analysis
4.5 A Comparison of Classification Methods
4.6 Footnotes

---

1 Logistic Regression

1.1 An Overview of Classification

- In ***classification*** we consider paired data $(\mathbf{X}, Y)$, where Y is a ***qualitative variable***, that is, a finite random variable.
- The values Y takes are called ***classes***

1.2 Why Not Linear Regression?

Because a linear regression model implies an ordering on the values of the response and in general there is no natural ordering on the values of a qualitative variable

1.3 Logistic Regression

1.3.1 The Logistic Model

- Consider a quantitiative predictor X binary response variable $Y \in 0, 1$
- We want to model the conditional probability of $Y = 1$ given X

$$P(X) := P(Y = 1|X)$$

- We model $P(X)$ with the ***logistic function***

$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

- The logistic model can be considered a linear model for the ***log-odds*** or ***logit***

$$\log\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 X$$

1.3.2 Estimating the Regression Coefficients

- The likelihood function for the logistic regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \prod_{i=1}^{n} p(x_i) \\ &= \prod_{i:y_i=1} p(x_i) \prod_{i:y_i=0} (1 - p(x_i)) \end{aligned}$$

- The maximum likelihood estimate (MLE) for the regression parameter is

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\operatorname{argmax}}\, \ell(\boldsymbol{\beta})$$

- There isn't a closed form solution for $\hat{\boldsymbol{\beta}}$ so it must be found using numerical methods

1.3.3 Making Predictions

- The MLE $\hat{\boldsymbol{\beta}}$ results in an estimate for the conditional probability $\hat{P}(X)$ which can be used to predict the class Y

1.3.4 Multiple Logistic Regression

- Multiple logistic regression considers the case of multiple predictors $\mathbf{X} = (X_1, \ldots, X_p)^\top$.
- If we write the predictors as $\mathbf{X} = (1, X_1, \ldots, X_p)^\top$, and the parameter $(\beta) = (\beta_0, \ldots, \beta_p)^\top$ then multiple logistic regression models

$$p(X) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{X})}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{X})}$$