Poisson regression, LDA and QDA

MSBA7002 Business Statistics Tutorial 3

ZHAO Yuwei

The University of Hong Kong

zhaoyw16@hku.hk

Table of Contents

Poisson regression

Suppose a random variable Y takes on nonnegative discrete values 0,1,2.... If Y follows a Poisson distribution with mean $\lambda(x_1, ..., x_p)$, the density function of Y is

$$Pr(Y = k) = \frac{exp(-\lambda(x_1, ..., x_p))\lambda(x_1, ..., x_p)^k}{k!} \quad \text{for k=0,1,...} \tag{1}$$

Poisson regression

For a Poisson regression, we assume that

$$log(\lambda(x_1, ..., x_p)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \tag{2}$$

because $\lambda(x_1, ..., x_p)$ is the mean of Y and has to be positive.

Poisson regression

Assume that for observation i, we have $(x_{i,1}, ..., x_{i,p})$ and y_i, the likelihood function is

$$L = \Pi_{i=1}^{n} \frac{exp(-\lambda(x_{i,1}, ..., x_{i,p}))\lambda(x_{i,1}, ..., x_{i,p})^{y_i}}{y_i!}. \tag{3}$$

In Poisson regression, we seek to maximize the likelihood function.

Poisson regression

Regression model	Linear	logistic	Poisson
Distribution	Gaussian	binomial	Poisson
Link function	$\mu = x'\beta$	$log(\frac{p}{1-p}) = x'\beta$	$log(\lambda) = x'\beta$

Table of Contents

Linear Discriminant Analysis

1. Discriminant Analysis
2. Linear Discriminant Anlysis
3. Quadratic Discriminat Analysis

Linear Discriminant Analysis

Discriminat Analysis

Step 1. Estimate the probability distribution for each class

$$\hat{f}_k(\mathbf{X}) = \hat{f}_k(X_1, \ldots, X_p), \quad k = 1, 2, \ldots, K$$

Step 2. Calculate the conditional probability of each observation for each class

$$\{\hat{p}(\mathbf{x_i}|\mathbf{y} = \mathbf{1}), \ldots, \hat{p}(\mathbf{x_i}|\mathbf{y} = \mathbf{K})\}, \quad i = 1, 2, \ldots, N$$

Step 3. Calculate the (posterior) probability of each observation for each class

$$\{\hat{p}(\mathbf{y} = \mathbf{1}|\mathbf{x_i}), \ldots, \hat{p}(\mathbf{y} = \mathbf{K}|\mathbf{x_i})\}, \quad i = 1, \ldots, N$$

Step 4. Predict the class based on estimated probability

$$k_i = \arg \max_{y \in \{1,\ldots,K\}} \{\hat{p}(\mathbf{y} = \mathbf{1}|\mathbf{x_i}), \ldots, \hat{p}(\mathbf{y} = \mathbf{K}|\mathbf{x_i})\}, \quad i = 1, \ldots, N$$

Linear Discriminant Analysis

Discriminat Analysis

Step 1

Estimate the probability distribution for each class

$$\hat{f}_k(\mathbf{X}) = \hat{f}_k(X_1, \ldots, X_p), \quad k = 1, 2, \ldots, K$$

- We need to assume the distribution of $(X_1, \ldots, X_p)$ for each group
  - Multivariate normal distribution

    LDA: assume the same covariance matrix for each group

    QDA: assume a different covariance matrix for each group
  - Other distributions

Linear Discriminant Analysis

Discriminat Analysis

Step 2

Calculate the conditional probability

$$\{\hat{p}(\mathbf{x_i}|\mathbf{y}=\mathbf{1}),\ldots,\hat{p}(\mathbf{x_i}|\mathbf{y}=\mathbf{K})\},\quad i=1,2,\ldots,N$$

- Suppose obsevation i belong to a given class
- if observation i belongs to class $k(k=1,2,\ldots,K)$,

$$\hat{p}(\mathbf{x_i}|y=k)=\hat{f}_1(\mathbf{x}_i)=\hat{f}_k(x_{i1},\ldots,x_{ip})$$

Linear Discriminant Analysis

Discriminat Analysis

Step 3

Calculate the (posterior) probability of each observation

$$\{\hat{p}(\mathbf{y}=\mathbf{1}|\mathbf{x_i}),\dots,\hat{p}(\mathbf{y}=\mathbf{K}|\mathbf{x_i})\},\quad i=1,\dots,N$$

- it is a Bayesian method
- we need to add prior probability of class $\{\pi_1,\pi_2,\dots,\pi_K\}$

$$\hat{p}(y=1|\mathbf{x_i})=\frac{\pi_1\hat{p}(\mathbf{x_i}|y=1)}{\sum_{k=1}^K\pi_k\hat{p}(\mathbf{x_i}|y=k)}$$

$$\vdots$$

$$\hat{p}(y=K|\mathbf{x_i})=\frac{\pi_K\hat{p}(\mathbf{x_i}|y=K)}{\sum_{k=1}^K\pi_k\hat{p}(\mathbf{x_i}|y=k)}$$

Linear Discriminant Analysis

Discriminat Analysis

Step 3

Calculate the (posterior) probability of each observation

$$\{\hat{p}(\mathbf{y} = \mathbf{1}|\mathbf{x_i}), \ldots, \hat{p}(\mathbf{y} = \mathbf{K}|\mathbf{x_i})\}, \quad i = 1, \ldots, N$$

- Prior Probability of Class $\{\pi_1, \pi_2, \ldots, \pi_K\}$
  - Prior information
  - Without any prior information
    - i Use proportation of each class as the prior

$$\pi_k = \frac{n_k}{\sum_{j=1}^{K} n_j}$$

    - ii Use equal prior probability

$$\pi_1 = \pi_2 = \ldots = \pi_K = 1/K$$

    - iii Other mehods

Linear Discriminant Analysis

Discriminat Analysis

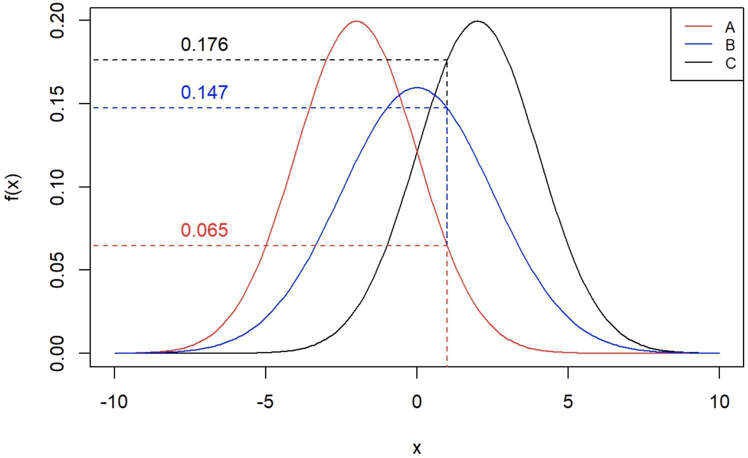
Step 4

Predict the class based on estimated probability

$$k_i = \arg\max_{k\in\{1,\ldots,K\}} \{\hat{p}(\mathbf{k}=\mathbf{1}|\mathbf{x_i}),\ldots,\hat{p}(\mathbf{k}=\mathbf{K}|\mathbf{x_i})\}, \quad i=1,\ldots,N$$

Linear Discriminant Analysis

Discriminat Analysis



Linear Discriminant Analysis

Discriminat Analysis

y	π_k	$p(\mathbf{x}\|y=k)$	$\pi_k p(\mathbf{x}\|y=k)$	$\frac{\pi_k p(\mathbf{x}\|y=k)}{\sum_{j=1}^{K} \pi_k p(\mathbf{x}\|y=j)}$	$\hat{y}$
A	0.6	0.065	0.039	0.3764479	✓
B	0.2	0.147	0.0294	0.2837838	
C	0.2	0.176	0.0352	0.3397683	
Σ			0.1036	1.0	

Linear Discriminant Analysis

Linear Discriminat Analysis

Three Hypotheses

i. $X_1, \ldots, X_p$ follow Mulitivariate Guassian Distribution
ii. Homoscedasticity assumption:
The class covariance matrices are identical

$$\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$$

Linear Discriminant Analysis

Linear Discriminat Analysis

Two features

- Easy to estimate probability function

$$f_k(x_1, \ldots, x_p) = \prod_{j=1}^{p} f_k(x_j)$$

$$f_k(x_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}(\frac{x_j - \mu_{jk}}{\sigma_j})^2}$$

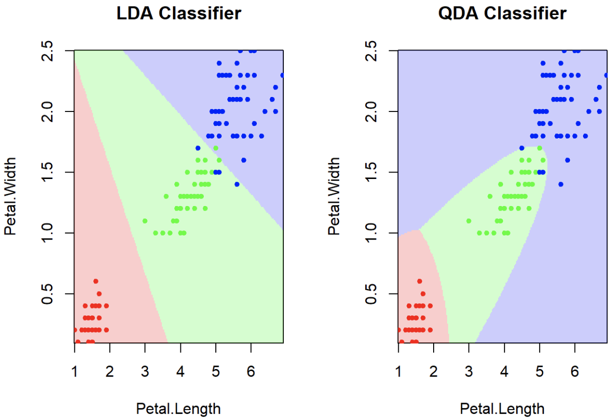
$$\ln(f_k(x_j)) = -\frac{1}{2}[(\frac{x_j - \mu_{jk}}{\sigma_j})^2 + \ln(2\pi) + \ln(\sigma_j^2))]$$

$$\frac{\partial f_k}{\partial \mu_{jk}} = 0, \quad \frac{\partial f_k}{\partial \sigma_j^2} = 0$$

Linear Discriminant Analysis

Linear Discriminat Analysis

ii Provide a linear Bayes Decision Boundary



Linear Discriminant Analysis

Linear Discriminat Analysis

ii Provide a linear Bayes Decision Boundary

- The rotation is meanful !
- Let Y "supervise" the rotation
  $\Rightarrow$ **Discriminant Variables**