

统计学习笔记

Stanley Wu
GmaxInfo
bleedtodry@hotmail.com

Contents

1. 线性回归	2
2. 分类算法	2
2.1. Logistic Regression	1
2.2. Linear Discriminant Analysis	1
3. 重抽样	2
3.1. Cross-Validation	2
3.1.1. Leave-One-Out Cross-Validation	2
3.1.2. k-Folder Cross-Validation	2

1.

2.

2.1. Logistic Regression

基本模型:

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

Log-odds logit:

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X \quad (2)$$

最大似然参数估计:

$$\zeta(\beta_0, \beta_1) = \prod_{i: y_i=1} p(x_i) \prod_{i: y_i=0} (1 - p(x_i)) \quad (3)$$

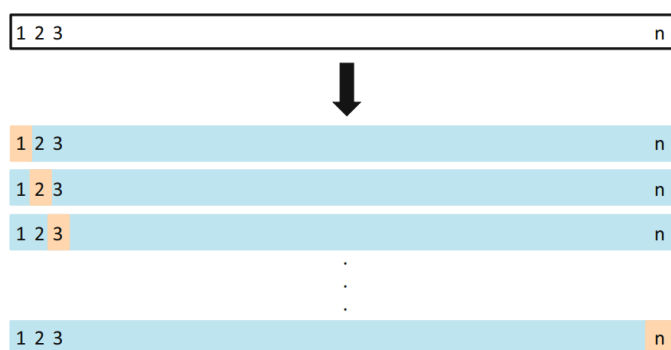
2.2. Linear Discriminant Analysis

Logistic Regression适用于二元离散回归, 当因变量大于2个时, 使用LDA

3.

3.1. Cross-Validation

3.1.1. Leave-One-Out Cross-Validation



$$CV = \frac{1}{n} \sum_{i=1}^n MSE_i \quad (4)$$

3.1.2. k-Folder Cross-Validation

$$CV_{(K)} = \sum_{k=1}^K \frac{n_k}{n} MSE_k \quad (5)$$

L00CV是k-Folder CV的一种特殊情况，即 $k = n$ 。