

Data+Science+??

March 23, 2018

1 Data Mining

KDD(Knowledge Discovery in Database) , .

1) (Exploration) , , 2) (Modification) 3) (Modeling) (Assessment) .

- (Cluster Analysis)
- (Link Analysis)
- (Discrimination Analysis)
- (Association Rule)
- (Decision Tree)
- (Neural Network)
- OLAP(On-Line Analytic Processing)

1.0.1 1. Logit Regression

() , .

, .

$$\log \frac{p(y = 1|x_1, \dots, x_p)}{1 - p(y = 1|x_1, \dots, x_p)} = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

\$x_1, x_2, \dots, x_p\$, .
(Logistic Discrimination) .

$$\hat{p}(y = 1|x_1, \dots, x_p) = \frac{\exp(\hat{a} + \hat{b}_1 x_1 + \dots + \hat{b}_p x_p)}{1 + \exp(\hat{a} + \hat{b}_1 x_1 + \dots + \hat{b}_p x_p)}$$

, 0 1 , (Cutoff value) .
(Odds Ratio) , $x_i + 1$.

$$\frac{\exp(\alpha + \beta_1 x_1 + \dots + \beta_i(x_i + 1) + \dots + \beta_p x_p)}{\exp(\alpha + \beta_1 x_1 + \dots + \beta_i(x_i) + \dots + \beta_p x_p)} = \exp(\beta_i)$$

1 x_i , 1 .