Prediction

Linear Regression

Common data mining models

- Classification
- Clustering
- Association Rule Discovery
- Prediction

Prediction

- Predict a value of a given continuous valued variable based on the values of other variables.

- Assuming a linear or nonlinear model of dependency.

Linear Regression

- This is the simplest form of linear regression, and it involves only one independent variable and one dependent variable. The equation for simple linear regression is:

$$y = \beta_0 + \beta_1 X$$

where:

- Y is the dependent variable
- X is the independent variable
- β0 is the intercept
- β1 is the slope

Linear Regression

- Least Square method

- Straight line $y = w_0 + w_1 x.$

$$w_1 = \frac{\sum_{i=1}^{|D|}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|}(x_i - \bar{x})^2}$$

$$w_0 = \bar{y} - w_1 \bar{x}$$

Find the salary for 10 years of experience

x years experience	y salary (in \$1000s)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

$$w_1 = \frac{(3-9.1)(30-55.4)+(8-9.1)(57-55.4)+\cdots+(16-9.1)(83-55.4)}{(3-9.1)^2+(8-9.1)^2+\cdots+(16-9.1)^2} = 3.5$$

$$w_0 = 55.4-(3.5)(9.1) = 23.6$$

Least squares line is estimated by
$y = 23.6 + 3.5x$.

Predict the salary of a college graduate with, 10 years of experience is $58,600.

Practice- Find the dollars for 50 shops

Shops	Dollars(millions)
22	64
25	74
29	82
35	90
38	100
42	120
46	120
52	142
65	180
88	230

Reference:

Data Mining: Concepts and Techniques , Jiawei Han and Micheline Kamber