CS685: Data Mining
Regression

Arnab Bhattacharya
arnabb@cse.iitk.ac.in

Computer Science and Engineering,
Indian Institute of Technology, Kanpur
http://web.cse.iitk.ac.in/~cs685/

1st semester, 2021-22
Mon 1030-1200 (online)

Regression

- Regression is mostly done for prediction
- Can also be viewed as a classifier but with a continuum of classes
- Describe how the response variable Y is generated given a set of k predictors, denoted by X

Regression

- Regression is mostly done for prediction
- Can also be viewed as a classifier but with a continuum of classes
- Describe how the response variable Y is generated given a set of k predictors, denoted by X
- General form of regression model is

$$Y = f(X) + \varepsilon$$

- Function f can be chosen using domain knowledge
- ε encodes the *error* terms associated with each observation
- Can predict missing values

Regression

- Regression is mostly done for prediction
- Can also be viewed as a classifier but with a continuum of classes
- Describe how the response variable Y is generated given a set of k predictors, denoted by X
- General form of regression model is

$$Y = f(X) + \varepsilon$$

- Function f can be chosen using domain knowledge
- ε encodes the *error* terms associated with each observation
- Can predict missing values
- Two main forms, linear regression and logistic regression

Linear Regression

- Function f is linear
- Response variable depends only *linearly* on the predictors
- General form of linear regression model is

$$Y = X \cdot W + \varepsilon$$

- W are the regression coefficients or *weights* on the predictors
- There are n observations
- Sizes of matrices

Linear Regression

- Function f is linear
- Response variable depends only *linearly* on the predictors
- General form of linear regression model is

$$Y = X \cdot W + \varepsilon$$

- W are the regression coefficients or *weights* on the predictors
- There are n observations
- Sizes of matrices
  - $Y : n \times 1$
  - $X : n \times k$
  - $W : k \times 1$
  - $\varepsilon : n \times 1$

Linear Regression

- Function f is linear
- Response variable depends only *linearly* on the predictors
- General form of linear regression model is

$$Y = X \cdot W + \varepsilon$$

- W are the regression coefficients or *weights* on the predictors
- There are n observations
- Sizes of matrices
  - $Y : n \times 1$
  - $X : n \times k$
  - $W : k \times 1$
  - $\varepsilon : n \times 1$
- Estimate W that minimizes the *square* of the error terms $\varepsilon^T \varepsilon$
- Solved using ordinary least squares (OLE) method

Linear Regression

- Function f is linear
- Response variable depends only *linearly* on the predictors
- General form of linear regression model is

$$Y = X \cdot W + \varepsilon$$

- W are the regression coefficients or *weights* on the predictors
- There are n observations
- Sizes of matrices
  - $Y : n \times 1$
  - $X : n \times k$
  - $W : k \times 1$
  - $\varepsilon : n \times 1$
- Estimate W that minimizes the *square* of the error terms $\varepsilon^T \varepsilon$
- Solved using ordinary least squares (OLE) method
- The least squares estimate $\widehat{W}$ is

$$\widehat{W} = (X^T X)^{-1} X^T Y$$

Logistic Regression

- Function f is the logistic function or sigmoid function
- General form of logistic regression model is

$$y = \frac{1}{1 + e^{-z}} + \varepsilon \text{ where } z = w \cdot x$$

- w are the regression coefficients
- z is linear in x
- Response y is a sigmoid function of linear combination of x_i
- y is constrained to be within 0 and 1

Logistic Regression

- Function f is the logistic function or sigmoid function
- General form of logistic regression model is

$$y = \frac{1}{1 + e^{-z}} + \varepsilon \text{ where } z = w \cdot x$$

- w are the regression coefficients
- z is linear in x
- Response y is a sigmoid function of linear combination of x_i
- y is constrained to be within 0 and 1
- Solved using gradient descent method
- No nice analytical solution

Logistic Regression

- Function f is the logistic function or sigmoid function
- General form of logistic regression model is

$$y = \frac{1}{1 + e^{-z}} + \varepsilon \text{ where } z = w \cdot x$$

- w are the regression coefficients
- z is linear in x
- Response y is a sigmoid function of linear combination of x_i
- y is constrained to be within 0 and 1
- Solved using gradient descent method
- No nice analytical solution
- Essentially a single layer perceptron/ANN with sigmoid activation function