Linear Regression

COGNIBOT

AI meets Industry

- Once we've acquired data with multiple variables, one very important question is how the variables are related.
- Regression is a set of techniques for estimating relationships

COGNIBOT
AI meets Industry

Fitting curves to bivariate data

- Bivariate data (x1,y1), (x2,y2), ..., (xn,yn)
- Model:
    - $Yi = f(xi) + Ei$ (where f(x) is some function, Ei random error)
- Total squared error:

$$\sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

COGNIBOT
AI meets Industry

- Model allows us to predict the value of y for any given value of x
    - X is called the independent variable
    - Y is the dependent variable

COGNIBOT

AI meets Industry

Examples of f(x)

- Lines: $y = ax + b + E$
- Polynomials: $y = ax\text{^}2 + bx + c + E$
- Other: $y = asin(x) + b + E$

COGNIBOT
AI meets Industry

Simple linear regression

Find the best-fitting line

COGNIBOT

AI meets Industry

- Fit a line to the data
  - $yi = axi + b + E$ (E is of normal distribution)
  - Total squared error:

$$\sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

- Goal: Find the values of a and b that give the best fit
- Best fit: minimizes the total squared error

COGNIBOT
AI meets Industry

Linear Regression

Find the best-fitting polynomial

COGNIBOT

AI meets Industry

- Fit a parabola to the data
  - $yi = axi\text{^}2 + bxi + c + E$ (E is of normal distribution)
  - Total squared error:

$$\sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - ax_i^2 - bx_i - c)^2.$$

- Goal: Find the values of a, b, c that give the ‘best fitting parabola’.
- Best fit: minimizes the total squared error
- Can also fit higher order polynomials

COGNIBOT
AI meets Industry

- Parameters are linear:
  - $y = ax + b$
  - $y = ax^2 + bx + c$
- It is not because the curve being fit has to be a straight line–although this is the simplest and most common case.

COGNIBOT
AI meets Industry

Examples of f(x)

- Lines: $y = ax + b + E$
- Polynomials: $y = ax\text{^}2 + bx + c + E$
- Other: $y = asin(x) + b + E$

COGNIBOT
AI meets Industry

Measuring the fit

- TSS = ∑(yi—y)^2 = total sum of squares = total variation.
- RSS = ∑(yi—yˆi)^2 = residual sum of squares.
- RSS/TSS = unexplained fraction of the total error.
- R^2= 1—RSS/TSS is measure of goodness-of-fit
- R^2 is the fraction of the variance of y explained by the model.

COGNIBOT
AI meets Industry