The University of Queensland Australia | CREATE CHANGE

Machine Leaning

COMP4702/COMP7703

Prac 4

Amelia Qiu

https://github.com/Amelia-Tong/MachineLearning_COMP4702/blob/main/week4

Parametric vs Non-parametric Models

- **Parametric** :

- assumes population distribution
- fixed parameters

- **Non-parametric** :

- no distribution assumption
- no fixed parameters

Linear Regression

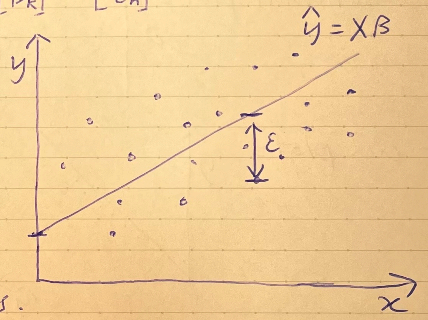
Linear Regression

- Target variable: numeric y
- Features: $x_1, x_2, \dots, x_k$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Or in matrix form: $Y = XB + \varepsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$



- β_0 (intercept)
  * The prediction when all features are 0.
- $\beta_{1,2,k}$ (coefficient)
  * can be imagined as slopes.
  * represents the expected change in y for a one-unit change in the corresponding x.
- ε (error) $= y - \hat{y}$
  * represents the difference between ~~true~~ the true value and the predicted value.

Ordenary Least Square (OLS) and Maximum Likelihood Estimate (MLE) provide a same close-form solution:

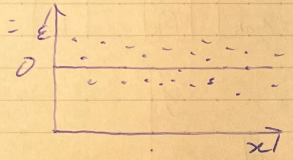
$$\hat{\beta} = (X'X)^{-1} X'y$$

under some assumptions:

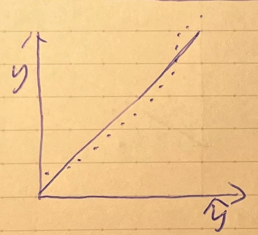
$$\varepsilon \sim IID, \quad E(\varepsilon \mid X) = 0, \quad X \sim N(0, \sigma^2)$$

To check these assumptions, we can use:

- Residual plot:



- QQ plot:



Polynomial Regression

Polynomial Regression

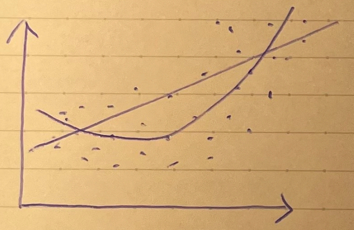
Extends linear regression by adding quaduatic, cubic, or higher-degree terms, so that the model can capture non-linear relationships.

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_d x^d + \varepsilon$ (only 1 variable x)

Having 2 features:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
$$+ \beta_3 x_1^2 + \beta_4 x_2^2$$
$$+ \beta_5 x_1 x_2 \text{ (interaction term)}$$

capture the effect of combining 2 or more features.



Logistic Regression

Logistic Regression

$$\underline{Y} = \underline{X\beta + \varepsilon}$$

However, $X\beta$ predict numerical output, just as in linear regression.

For classification problem, our response Y is categorical.

Thus, instead of directly predicting the result, we are now predicting the log-odd probability:

$$\log\left(\frac{p}{1-p}\right) = X\beta + \varepsilon$$

Logistic function $P = e^{\eta}/1+e^{\eta}$ transforms the log-odd probability to probability: