Overfitting and Underfitting

TRAIN	Very Good Accuracy [Low Bias]
TEST	Very Good Accuracy [Low Variance]

⇑

Generalized model

Very Good Accuracy [Low Bias]
Bad Accuracy [High Variance]

⇓

Model is overfitting

TRAIN	Accuracy is low [High Bias]
TEST	Accuracy is low [High Variance]

⇓

Model is underfitting

Ordinary least Square (OLS)

$$h_\theta(x) = \beta_0 + \beta_1 x$$

$$S(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\text{Intercept} = \beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})}$$

Polynomial Regression

Simple Polynomial Regression

degree 0 $\Rightarrow$ $h_\theta(x) = \beta_0 x^0$ $\longrightarrow$ Constant value

degree 1 $\Rightarrow$ $h_\theta(x) = \beta_0 x^0 + \beta_1 x^1$ $\rightarrow$ Simple linear Regression

degree 2 $\Rightarrow$ $h_\theta(x) = \beta_0 x^0 + \beta_1 x^1 + \beta_2 x^2$

:

degree n $\Rightarrow$ $h_\theta(x) = \beta_0 x^0 + \beta_1 x^1 + \dots + \beta_n x^n$

Polynomial Regression (with multiple independent features)

degree 1 $\Rightarrow$ $h_\theta(x) = \beta_1 x_1 + \beta_2 x_2$

degree 2 $\Rightarrow$ $h_\theta(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2$