$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b} \underbrace{\frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)^2}_{\text{mean squared error cost}} + \underbrace{\frac{\lambda}{2m}\sum_{j=1}^{n} W_j^2}_{\text{regularization term}}$$

Two goals of this cost function :

① mean squared error cost term ⇒ learning algorithm to fit the training data well by minimizing the squared differences between the predictions and the targets

② regularization term ⇒ keep the parameters (W_j) small which will tend to reduce overfitting

lambda (λ) value 를 적절한 값으로 설정하는 것이 전체 cost function에 중요한 요인으로 작용

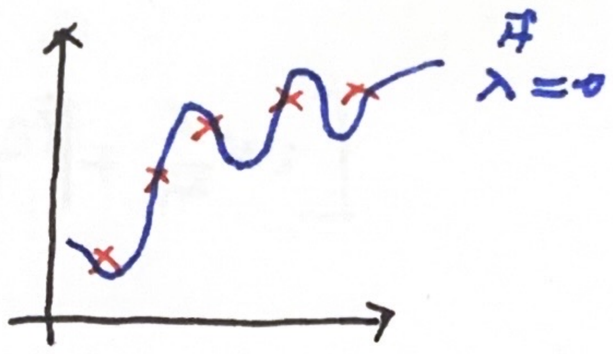
㉮ $\lambda = 0$

⇒ regularization term = 0

= "no penalty"

= "overfitting"

⇒



㉯ $\lambda = 10^{10}$ (enormous value)

⇒ To minimize regularization term, all the parameteres W_j should be very close to zero (≈ 0)

⇒ $f_{\vec{w},b}(\vec{x}) = \cancel{W_1 x} + \cancel{W_2 x^2} + \cancel{W_3 x^3} + \cdots + b$

$\approx 0 \quad \approx 0 \quad \approx 0$

$= b$

⇑