< Regularization and bias / variance >

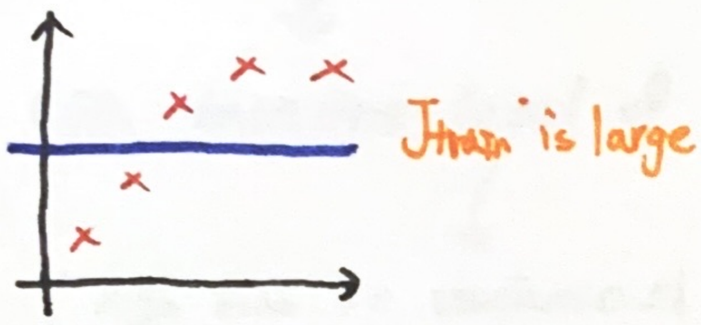
- choice of regularization parameter lambda(λ) affects the bias and variance

ex)

Model: $f_{\vec{w},b}(x) = w_1x + w_2x^2 + w_3x^3 + w_4x^4 + b$

Cost: $J(\vec{w},b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$

i) Large λ



$\lambda = 10,000$

$\Downarrow$

$w_1 \approx 0,\ w_2 \approx 0 \cdots$

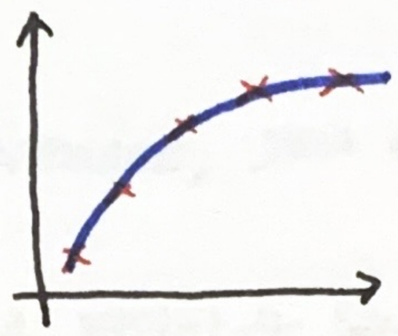
$\Downarrow$

$f_{\vec{w},b}(\vec{x}) \approx b$

$\Downarrow$

"High bias"
(underfit)

ii) Intermediate λ



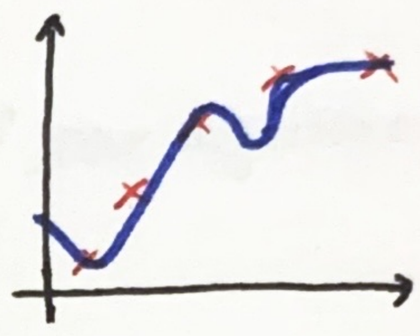
Jtrain is low

Jcv is low

$\Downarrow$

"Just right"

iii) Small λ



$\lambda = 0$

Jtrain is low

Jcv is high

$\Downarrow$

"High variance"
(overfit)

* Validation set을 이용하여 λ 선택

1. Try $\lambda = 0$ → $\min_{\vec{w},b} J(\vec{w},b)$ → $w^{<1>}, b^{<1>}$ → $J_{cv}(w^{<1>}, b^{<1>})$
2. Try $\lambda = 0.01$ → $\min_{\vec{w},b} J(\vec{w},b)$ → $w^{<2>}, b^{<2>}$ → $J_{cv}(w^{<2>}, b^{<2>})$
3. Try $\lambda = 0.02$

⋮ ⋮

12. Try $\lambda = 10$ → $\min_{\vec{w},b} J(\vec{w},b)$ → $w^{<12>}, b^{<12>}$ → $J_{cv}(w^{<12>}, b^{<12>})$

$\Downarrow$

J_{cv}가 가장 작을때가 $J_{cv}(w^{<5>}, b^{<5>})$라면 Pick

⇒ Report test error: $J_{test}(w^{<5>}, b^{<5>})$

* λ와 Jtrain / Jcv 상관관계