< Regularized Linear Regression >

* cost function

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

* Linear Regression model

$$f_{\vec{w},b}(\vec{x}) = \vec{w}\cdot\vec{x} + b$$

* gradient descent

repeat until convergence {

$$w_j := w_j - \alpha \boxed{\frac{\partial}{\partial w_j} J(\vec{w},b)}$$

$$b := b - \alpha \boxed{\frac{\partial}{\partial b} J(\vec{w},b)}$$

} simultaneous update

derivative ⟹

$$\frac{1}{m}\sum_{i=1}^{m}\left(\boxed{f_{\vec{w},b}(\vec{x}^{(i)})} - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m} w_j$$

$$\frac{1}{m}\sum_{i=1}^{m}\left(\boxed{f_{\vec{w},b}(\vec{x}^{(i)})} - y^{(i)}\right)$$

* don't have to regularize b (by convention)

⇓

- Implementing gradient descent

repeat until convergence {

$$w_j := w_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left[\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)}\right] + \frac{\lambda}{m} w_j\right]$$

$$b := b - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

} simultaneous update

$$w_j := 1\cdot w_j - \alpha\frac{\lambda}{m} w_j - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$= w_j\left(1 - \alpha\frac{\lambda}{m}\right) - \underbrace{\alpha\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}}_{\text{usual gradient descent update for unregularized}}$$

usual gradient descent update for unregularized ~~gradient descent~~

ex) $\alpha\frac{\lambda}{m} = 0.01\cdot\frac{1}{50} = 0.0002$

$\Rightarrow w_j\left(1-\alpha\frac{\lambda}{m}\right) = w_j(1-0.0002)$

$= w_j \times 0.9998$

every single update iteration

$(\alpha>0, \lambda>0, m>0)$

: multiplying w by a number of slightly less than 1

⇒ effect of shrinking the value of w_j little bit

||

"The way Regularization Works"