<Gradient Descent for Linear Regression>

① Linear Regression Model

$$f_{w,b}(x) = wx+b$$

② Cost function : Squared Error Cost Function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2$$

objective : $\underset{w,b}{\text{minimize}}\ J(w,b)$

③ Gradient Descent Algorithm

repeat until convergence {

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

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

] update w, b simultaneously

}

(calculate)

i) $\frac{\partial}{\partial w}J(w,b) = \frac{\partial}{\partial w}\frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2 = \frac{\partial}{\partial w}\ \Sigma$

(partial derivative)

$= \frac{1}{\cancel{2}m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)\cancel{2}x^{(i)} = \frac{1}{m}$

* Batch Gradient Descent

"Batch" : Each step of gradient descent uses all the training examples.

x size in feet²	y price in $1,000's
2104	400
1416	232
1534	315
852	178
⋮	⋮

$m = 47$

$\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2$

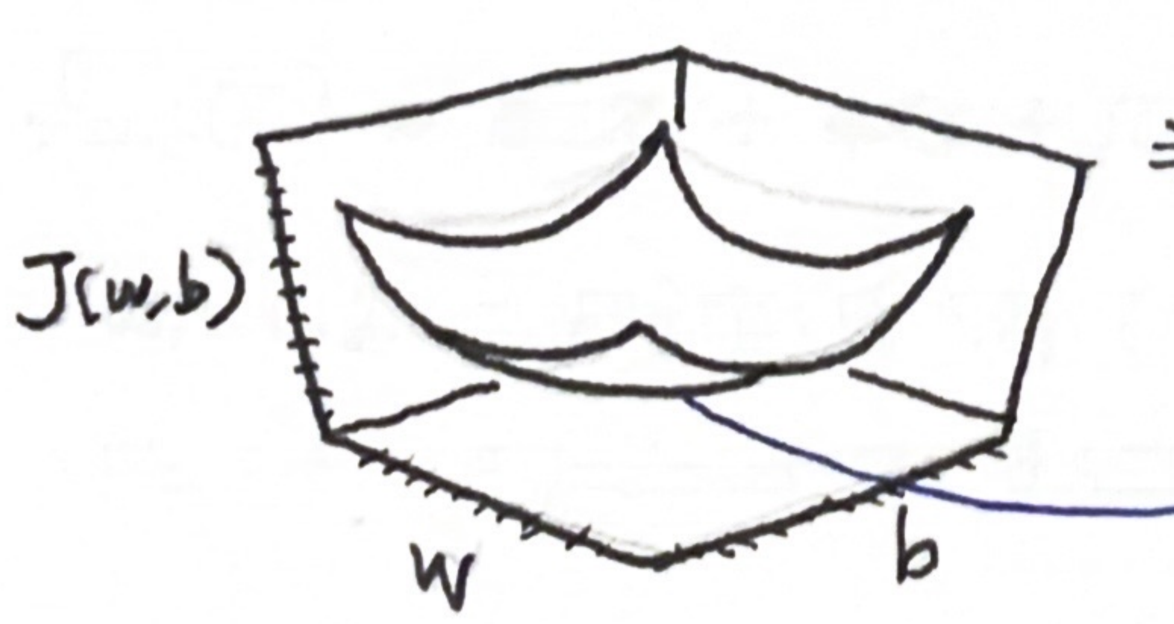
ii) $\frac{\partial}{\partial b}J(w,b) = \frac{\partial}{\partial b}\frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2 = \frac{\partial}{\partial b}\frac{1}{2m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)^2$

$= \frac{1}{\cancel{2}m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)\cancel{2} = \frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$

<Conclusion>

① 처음에 임의로 설정한 parameter (w,b)에서 출발하여 iteration이 거듭될수록 정확한 hypothesis(model)가 됨.

② Linear Regression의 cost function (Squared Error)은 convex function이므로 여러개의 local minimum이 존재할 수 없으며, local minimum이 곧 global minimum이기에 항상 global minimum에 수렴한다.



⇒ Squared Error cost function (convex function)

: bowl shape

local minimum = global minimum