Machine Learning Specialization

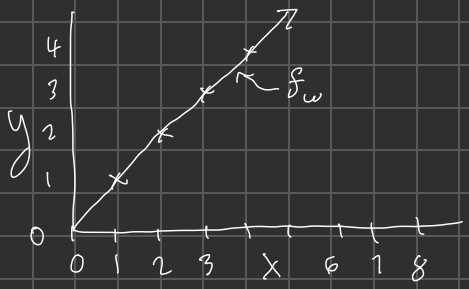
The Cost Function + Gradient Descent

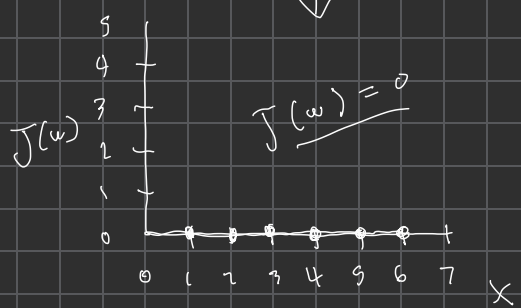
$f_w(x)$ = function of x, fixed w

$J(w)$ = function of w

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

$$= \frac{1}{2m} \sum_{i=1}^{m} \left(w x^{(i)} - y^{(i)} \right)^2 = \frac{1}{2m}(0^2 + 0^2 + 0^2)$$

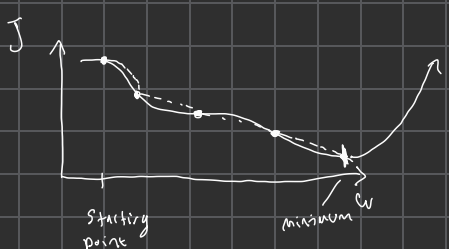


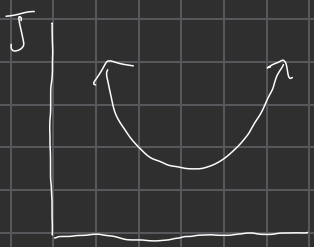


$\min_{w,b} J(w,b)$ Minimize the cost fn

$\min_{w_1, w_2 \ldots w_n, b} J(w_1, w_2, \ldots, w_n, b)$

- Start with some w, b, then iteratively change values to minimize the cost fn





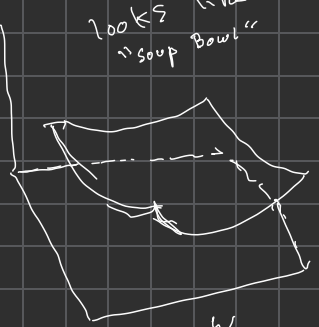
Main update formula: $w = w - \alpha \frac{d}{dw} J(w,b)$

$x, y \in \mathbb{R}$

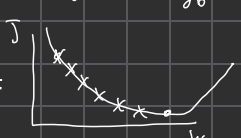
learning rate

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

$J(w,b)$



α is small: then lots of small updates



fail to converge and even diverge

α is big, lots of big updates