Recall:

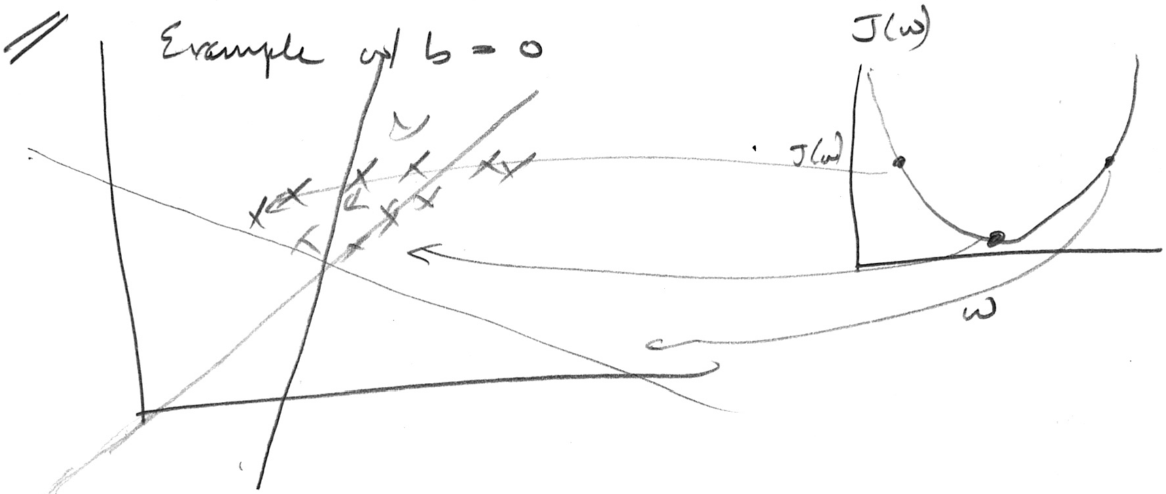
Model: $f_{w,b}(x) = wx + b$

Parameters: w, b

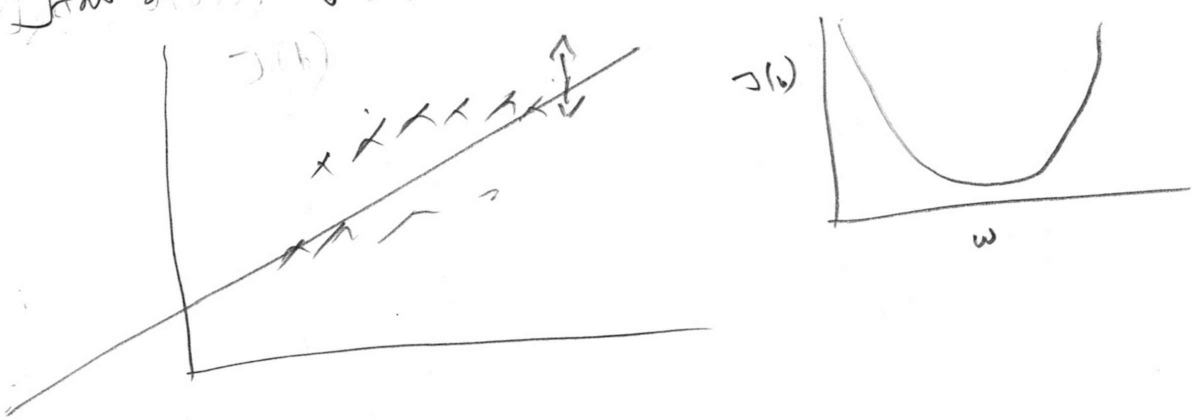
Cost Function: $J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} (\underbrace{f_{w,b}(x^{(i)})}_{\hat{y}^{(i)}} - y^{(i)})^2$

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

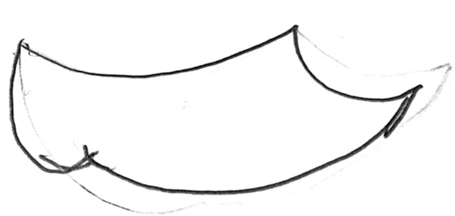
// Example w/ b = 0



How about if we fix w and scan b



⇒ J(w,b)



⇒ use graphing

Contour plot

We see that finding the minimum corresponds to model w/ smallest error on training data

$\Rightarrow$ Training = minimizing the cost function

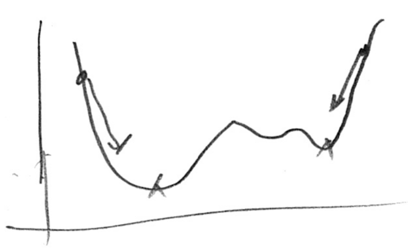
How? Gradient Decent (aka Newton's method)

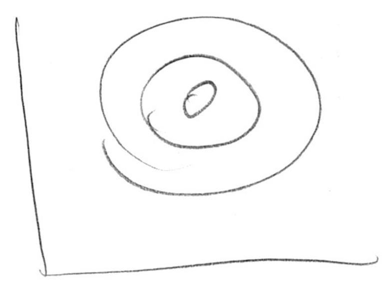
Take function $J(w,b)$ (or any func really)

want $\min_{w,b} J(w,b)$ $\min_{w_1,\ldots} J(w_1 - w_n)$

Strategy :
- Start w/ some w,b (e.g. set $w=0, b=0$)
- Keep changing w,b to reduce $J(w,b)$
- Stop when at or near minimum

Note that there could be multiple minima (not for MSE)





not math ... code
(equivalence) (assignment)

$$w_{k+1} = w_{k+1} - \alpha \frac{\partial}{\partial w} J(w_{k+1}, b)$$

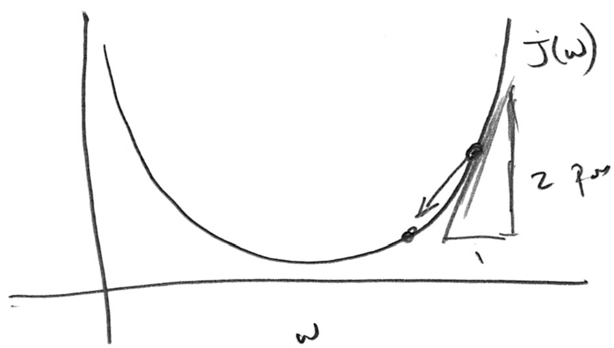
learning rate

$$b_{k+1} = b_{k+1} - \alpha \frac{\partial}{\partial b} J(w_k, b)$$

correct: update simultaneously

incorrect: update are at time
(but OK!)

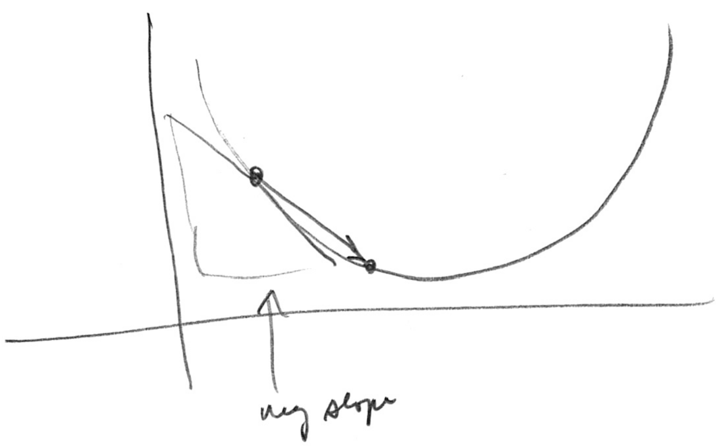
⇒ Repeat until Convergence



$$w_{k+1} = w_k - \alpha \underbrace{\frac{d}{dw} J(w)}_{\text{slope}}$$

this case ⇒ $w_{k+1} = w_k - \alpha$ (pos #)

⇒ move left



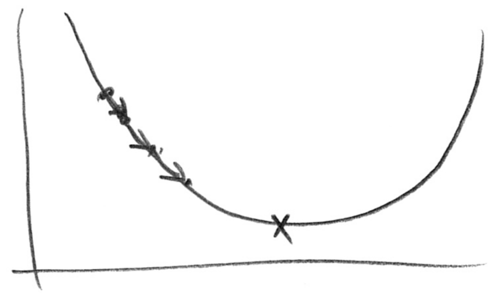
⇒ $w_{k+1} = w_k - \alpha$ (neg #)

⇒ move right

Learning Rate

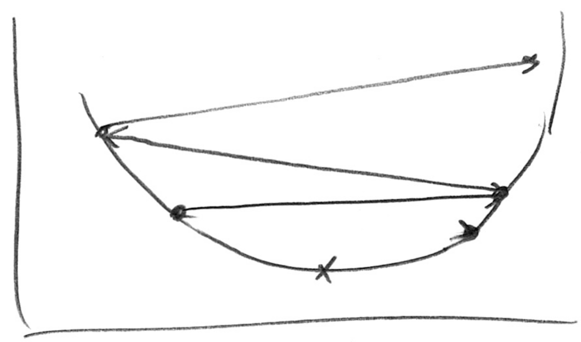
$$w_{k+1} = w_k - \alpha \frac{d}{dw} J(w)$$

small α :

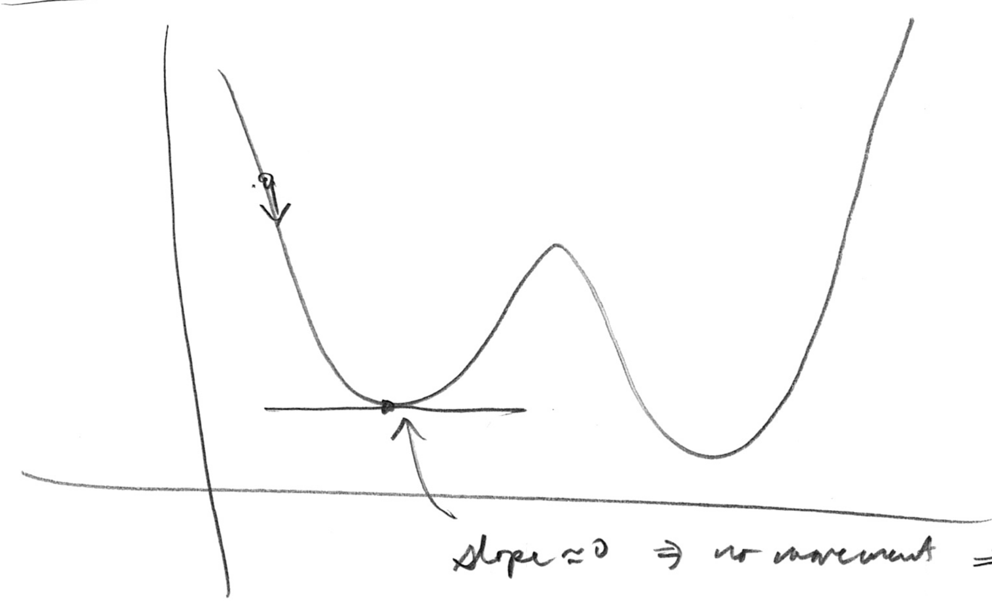


$\Rightarrow$ slow

large α :



$\Rightarrow$ overshoot
fail to
converge

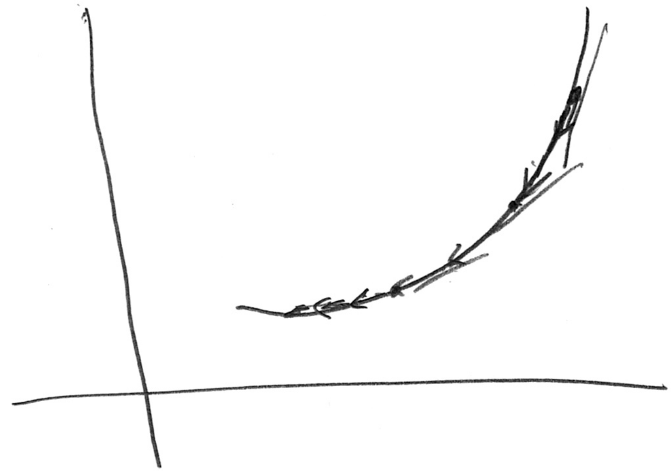


slope ≈ 0 $\Rightarrow$ no movement $\Rightarrow$ get stuck at
local minimum

Fixed learning rate

$$w_{k+1} = w_k - \alpha \frac{\partial}{\partial w} J(w)$$

as approach minimum,
slope decreases $\Rightarrow$
smaller steps



Gradient Descent for Linear Regression

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

Compute $\frac{\partial}{\partial w} J$ & $\frac{\partial}{\partial b} J$

$$\frac{\partial}{\partial w} J(w,b) = \frac{\partial}{\partial w}\frac{1}{2m}\sum_{i=1}^{m}\Big[\overbrace{(wx^{(i)}+b)}^{f_{w,b}(x^{(i)})} - y^{(i)}\Big]^2$$

$$= \cancel{2}\,\frac{1}{\cancel{2}m}\sum_{i=1}^{m}\Big[\underbrace{(wx^{(i)}+b)}_{f_{w,b}(x^{(i)})} - y^{(i)}\Big]x^{(i)}$$

$$\frac{\partial}{\partial b} J(w,b) = \cancel{\frac{\partial}{\partial b}}\ 2\ \frac{1}{2m}\sum_{i=1}^{m}\Big[(wx^{(i)}+b)^2 - y^{(i)}\Big]^2$$

$$\cancel{2}\,\frac{1}{\cancel{2}m}\sum\Big[(\qquad) - y^{(i)}\Big]$$

3 $\Rightarrow$

$$w_{k+1} = w_k - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b} \left(x^{(i)} - y^{(i)} \right) - y^{(i)} \right) x^{(i)}$$

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