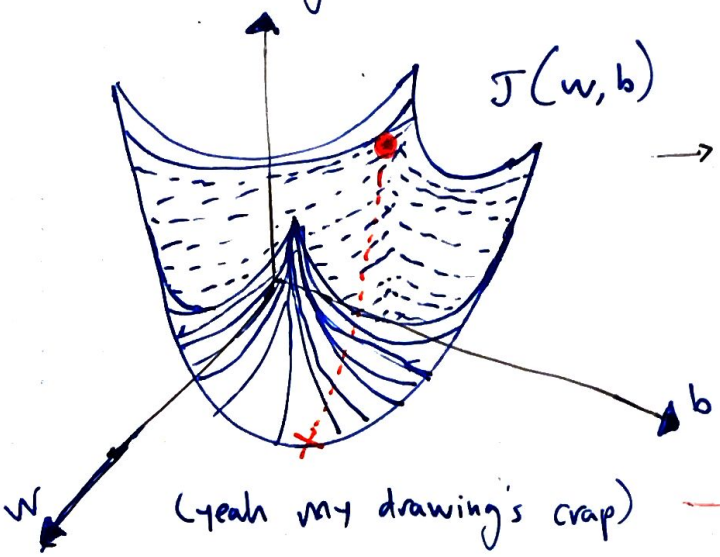
Gradient Descent:

Now let's actually try & use Gradient descent to learn params w & b!



→ want w, b to minimize J(w,b)

→ initialize w, b to some initial val

→ for log reg, anything works,
→ random → 0

(yeah my drawing's crap)

→ gradient descent starts at that initial pt. & takes a step in steepest downhill direction

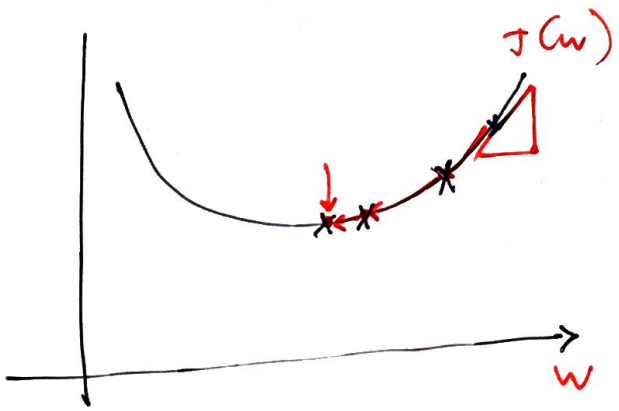
algo

Repeat {

$$w := w - \alpha \frac{dJ(w)}{dw}$$

(update, derivative; α: learning rate)

}



$$b := b - \alpha \frac{dJ(w,b)}{db}$$

$$w := w - \alpha \frac{(dJ(w,b))}{dw}$$

→ also can write as $\frac{\partial J(w,b)}{\partial w}$

* use ∂ if 2 or more params (partial deriv)
d if one

means the same thing mostly - or