Training Linear Regression

Gradient Decent

Outline

Have some function J(w,b)

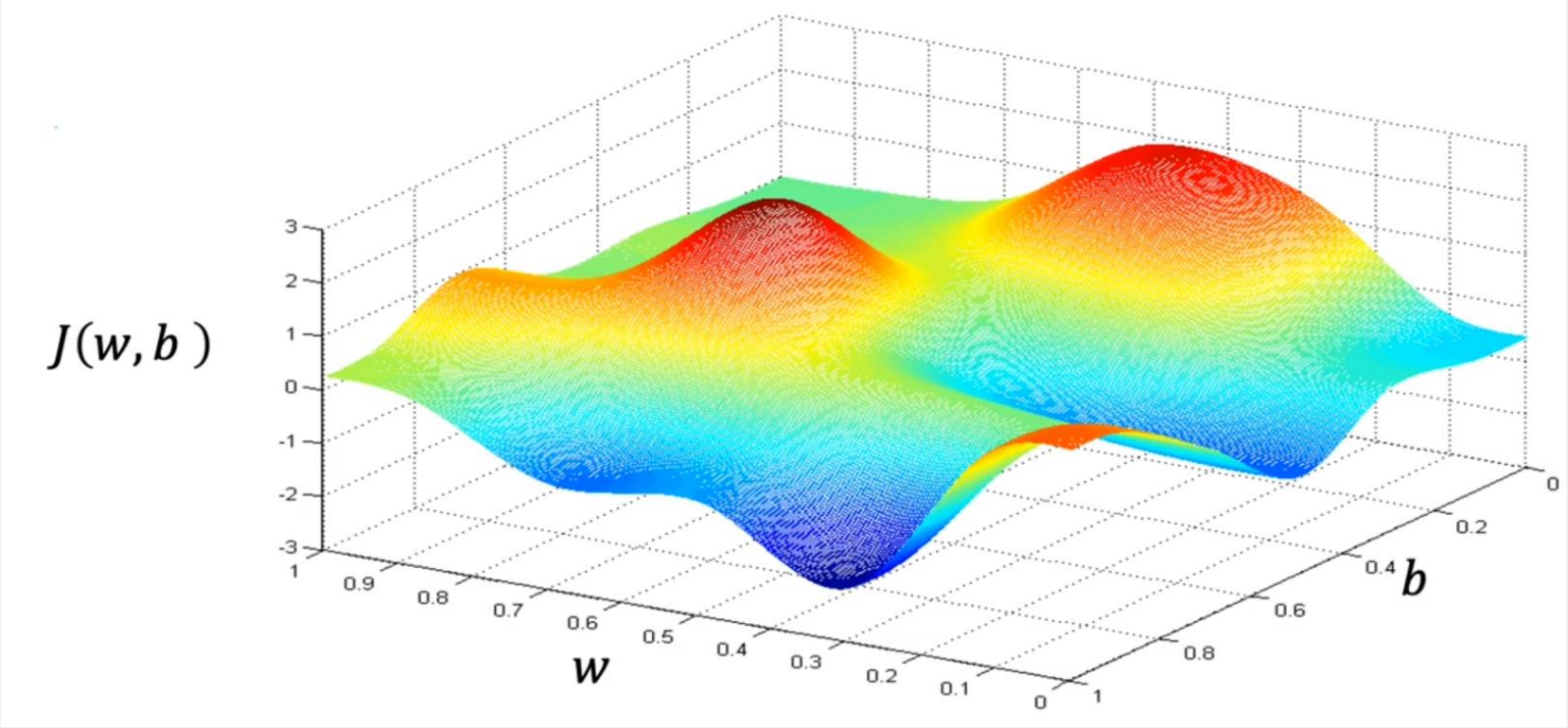
Want to minimize J(w,b) over parameters w & b

Start with any initial value of parameters w & b

Keep changing w & b to reduce J(w,b) until we settle at or near minimum

Function with more than one minimum?

Function with more than one minimum



Gradient decent algorithm

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

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

Simultaneous update

$$\text{temp } w := w - \alpha \frac{\partial J(w,b)}{\partial w} \quad -(1)$$

$$\text{temp } b := b - \alpha \frac{\partial J(w,b)}{\partial b} \quad -(2)$$

$$w = \text{temp } w \quad -(3)$$

$$b = \text{temp } b \quad -(4)$$

Incorrect

(1) → (3) → (2) → (4)

X

Simplified Gradient decent algorithm

repeat until convergence {

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

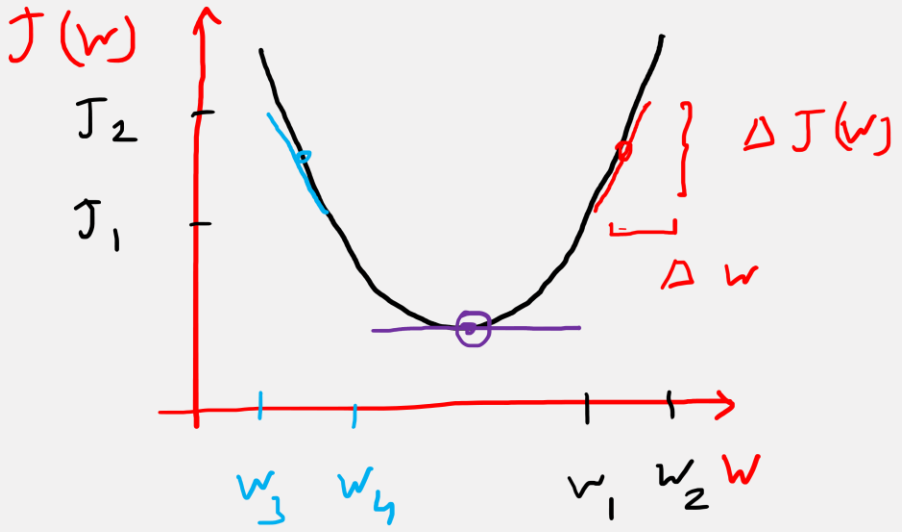
$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$J(w)$$

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

$$\min_w J(w)$$

Gradient decent algorithm intuition



$$\frac{\partial J(w)}{\partial w} = \frac{\Delta J(w)}{\Delta w} = \text{Positive}$$

$$\frac{\partial J(w)}{\partial w} = \text{Negative}$$

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

Learning rate α

If α is too small, then Gradient decent may converge very slowly.

If α is too large, then Gradient decent may overshoot resulting into diverging rather converging.