$$f(x : w_0 w_1) = w_0 + w_1 x$$

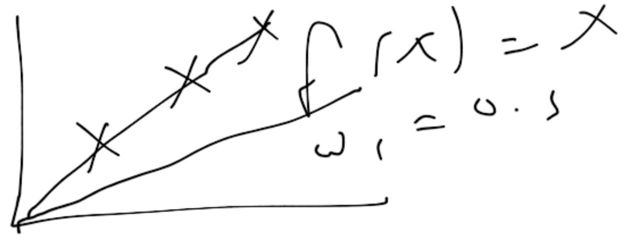
$$L = \frac{1}{N} \sum \left(f(x_n : w_0, w_1) - t_n \right)^2$$

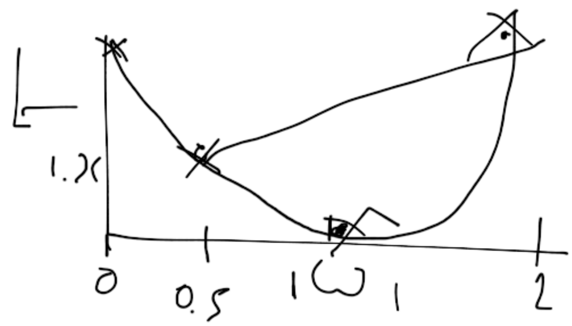
$w_0 = 0$

$(x_1 = 1, t_1 = 1$

$x_2 = 2 \quad t_2 = 2$

$x_3 = 3 \quad t_3 = 3$





$w_1 = 2$

$\frac{1}{3}\left((1)^2 + (2)^2 + 3^2 \right)$

$= \frac{14}{3} = 4.3$

Gradient descent

Start guess

$w_0^0 \quad w_1^0 \qquad L(w_0, w_1)$

Changes w_0, w_1 to reduce L

Alg

$$w_j^{new} = w_j^{old} - \alpha \frac{\partial L(w_0, w_1)}{\partial w_j}$$

Simult.

$$temp0 = w_0^{old} - \alpha \frac{\partial L}{\partial w_0}(w_0^{old}, w_1^{old})$$

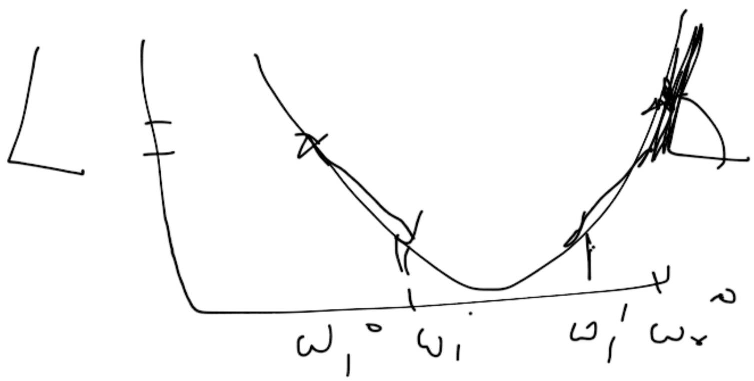
$$temp1 = w_1^{old} - \alpha \frac{\partial L}{\partial w_1}(w_0^{old}, w_1^{old})$$

$w_0^{new} = temp\,0$

$w_1^{new} = temp\,1$

Intuition

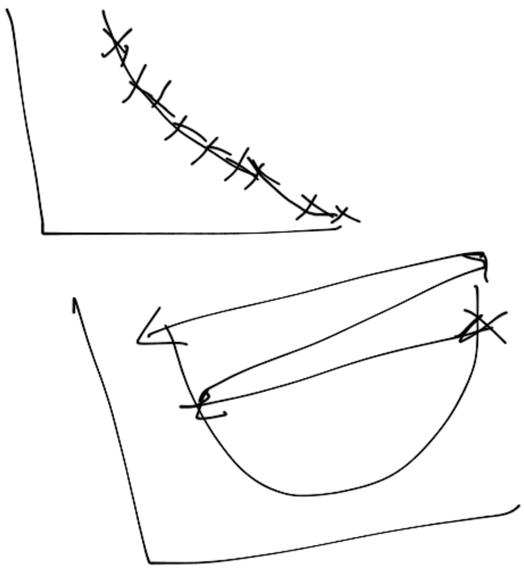
$$L = \frac{1}{N} \sum (w_1 x_n - t_n)^2$$



$$w_1' = w_1^0 - \alpha \frac{\partial L}{\partial w_1}$$

$w_1^0 - \alpha +$

$w_1^0 +$



α large.

$$\tilde{w_1} = w_1^{\circ} - \alpha \frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w}$$

$$f(x) = w_0 + w_1 x$$

$$L = \frac{1}{2N} \sum \left(f(x_n : w_0, w_1) - t_n \right)^2$$

$$\frac{\partial L}{\partial w_j} = \frac{1}{2N} \frac{\partial}{\partial w_1} \sum \left(w_0 + w_1 x - t_n \right)$$

$$\frac{\partial L}{\partial w_0} = \frac{1}{N} \sum \left(w_0 + w_1 x_n - t_n \right)$$

$$\frac{\partial L}{\partial w_1} = \frac{1}{N} \sum \left(\left(w_0 + w_1 x_n - t_n \right) x\, x_n \right)$$

$$\text{temp}_0 = w_0^{old} - \alpha \frac{\partial L}{\partial w_0}$$

$$\text{temp}_1 = w_1^{old} - \alpha \frac{\partial L}{\partial w_1}$$

Size, # be, # floors, Age

d = # features

x_m^i = i th feature x ...

$$X_2 = \begin{bmatrix} 1416 \\ 2 \\ 2 \\ 40 \end{bmatrix}$$

$$x_2^3 = 2$$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2^1 & x_2^2 & & x_2^d \\ \vdots & & & & \\ 1 & x_N^1 & x_N^2 & \cdots & x_N^d \end{bmatrix}$$

$$f = w^T x$$

$\sigma \cdot p$ mv V

Repeat

$$tmp_j = w_j^{old} - \alpha \frac{\partial L}{\partial w_j}$$

$$w_j^{new} = tmp_j$$

until convergence

$$L = \sum (f(x_i) - t_n)^2$$

$$\text{temp}_j = w_j^{old} - \alpha \sum \left(f(x_n ; w) - t_n \right) x_n^j$$

Linear modeling

$$t = f(x)$$

Examples $\begin{cases} wx \\ w_0 + w_1 x \\ w_0 + w_1 x^2 \end{cases}$

Good model

Loss function

$$L_n = (t_n - f(x_n ; w_0, w_1))^2$$

$$L_n \; | t_n - f(x_n ; w_0, w_1)|$$

$$\arg\min \frac{1}{N}\sum_{n=1}^{N} L_n(t_n, f(x_n : w_0, w_1))$$

$$L_n = \frac{1}{N}\sum \left(t_n - (w_0 + w_1 x_n)\right)^2$$

$$= \frac{1}{N}\sum w_1^2 x_n^2 + 2 w_1 x_n w_0 t_n + w_0^2 - 2 w_0 t_n + t_n^2$$

$$\frac{\partial L}{\partial w_0} = 0$$

$$w_0 = \frac{1}{N}\sum t_n - w_1 \frac{1}{N}\sum x_n$$

$$\hat{w}_0 = \bar{t} - w_1 \bar{x}$$

$$\frac{\partial L}{\partial w_1} = 2 w_1 \frac{1}{N} \left(\sum x_n^2 \right) + \frac{2}{N} \left(\sum x_n (w_0 - t_n) \right)$$

$$= 0$$

$$\hat{w}_1 = \frac{\frac{1}{N} \sum x_n t_n - \bar{t}\,\bar{x}}{\frac{1}{N} \sum x_n^2 - \bar{x}\,\bar{x}}$$

$$= \frac{\overline{xt} - \bar{x}\,\bar{t}}{\overline{x^2} - (\bar{x})^2}$$

$$L = \frac{1}{N} \sum_{n=1}^{N} (t_n - w^T x_n)^2$$

or

$$\frac{1}{N} (t - Xw)^T (t - Xw)$$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

$$t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}$$

$$t - Xw = \begin{bmatrix} t_1 - w_0 - w_1 x_1 \\ t_2 - w_0 - w_1 x_2 \\ \vdots \\ t_N - w_0 - w_1 x_N \end{bmatrix}$$

$$(Xw - t)^T (Xw - t)$$

$$= (w_0 + w_1 x_1 - t_1)^2 + (w_0 + w_1 x_2 - t_2)^2$$

$$\vdots$$

$$(w_0 + w_1 x_N - t_N)^2$$

$$L = \frac{1}{N}\left((Xw)^T - t^T\right)(Xw - t)$$

$$= \frac{1}{N}(Xw)^T Xw - \frac{1}{N} t^T X w - \frac{1}{N}(Xw)^T t + \frac{1}{N} t^T t$$

$$= \frac{1}{N} w^T X^T X w - \frac{2}{N} w^T X^T t + \frac{1}{N} t^T t$$

$$\frac{\partial L}{\partial w} = \begin{bmatrix} \frac{\partial L}{\partial w_0} \\ \frac{\partial L}{\partial w_1} \end{bmatrix}$$

$$\frac{\partial L}{w_0} = 2w_0 + 2w_1\bar{x} - 2\bar{t}$$

$$\frac{\partial L}{\partial w_1} = 2w_0\bar{x} + 2w_1\overline{x^2} - 2\overline{xt}$$

$$X^T X w = X^T t$$

$$(X^T X)^{-1}$$

$$J\ w = (X^T X)^{-1} X^T t$$

$$\hat{w} = (X^T X)^{-1} X^T t$$

$$f(x : w_0 w_1) = w_0 + w_1 x$$

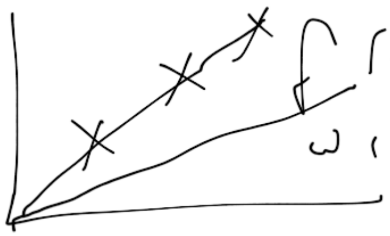
$$L = \frac{1}{N} \sum \left(f(x_n : w_0, w_1) - t_n \right)^2$$

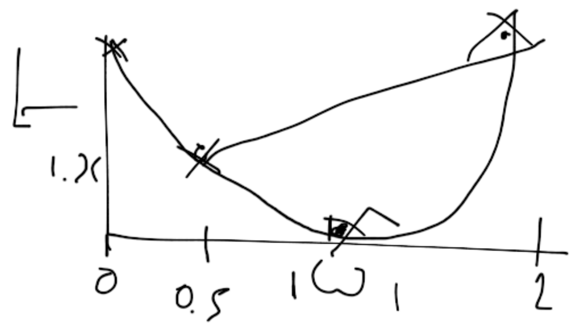
$w_0 = 0$

$(x_1 = 1, t_1 = 1$

$x_2 = 2 \quad t_2 = 2$

$x_3 = 3 \quad t_3 = 3$





$w_1 = 2$

$\frac{1}{3}\left((1)^2 + (2)^2 + 3^2 \right)$

$= \frac{14}{3} = 4.3$

Gradient descent

Start guess

w_0° w_1° $L(w_0, w_1)$

Changes w_0, w_1 to reduce L

Alg

$$w_j^{new} = w_j^{old} - \alpha \frac{\partial L(w_0, w_1)}{\partial w_j}$$

Simult.

$$temp_0 = w_0^{old} - \alpha \frac{\partial L}{\partial w_0}(w_0^{old}, w_1^{old})$$

$$temp_1 = w_1^{old} - \alpha \frac{\partial L}{\partial w_1}(w_0^{old}, w_1^{old})$$

$w_0^{new} = temp\,0$
$w_1^{new} = temp\,1$

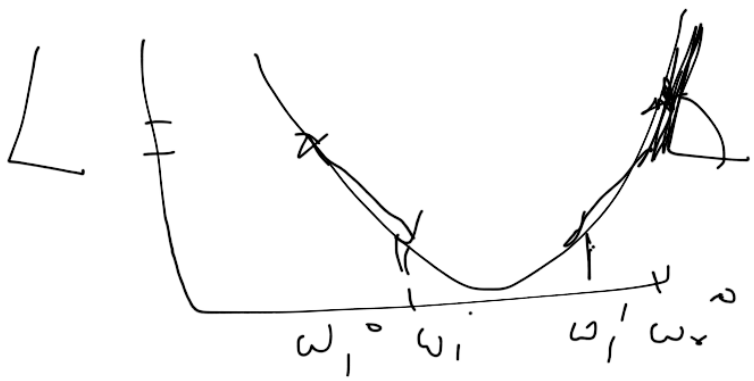
Intuition

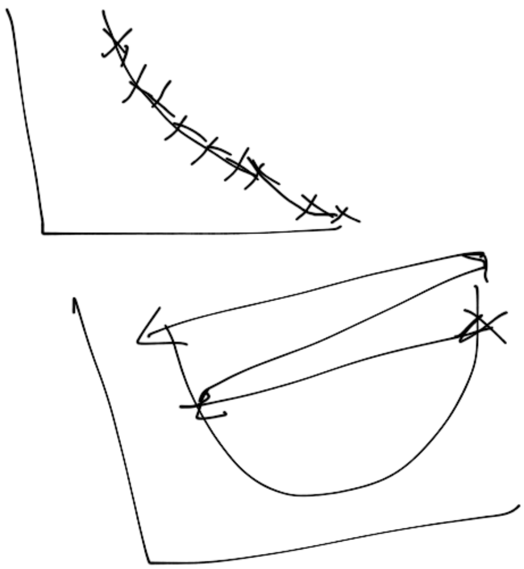
$$L = \frac{1}{N} \Sigma (w_1 x_n - t_n)^2$$

$$w_1^1 = w_1^0 - \alpha \frac{\partial L}{\partial w_1}$$

$w_1^0 - \alpha +$

$w_1^0 +$





α large.

$$\tilde{w_1} = w_1^{\circ} - \alpha \frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w}$$

$$f(x) = w_0 + w_1 x$$

$$L = \frac{1}{2N} \sum \left(f(x_n : w_0, w_1) - t_n \right)^2$$

$$\frac{\partial L}{\partial w_j} = \frac{1}{2N} \frac{\partial}{\partial w_1} \sum (w_0 + w_1 x - t_n$$

$$\frac{\partial L}{\partial w_0} = \frac{1}{N} \sum (w_0 + w_1 x_n - t_n)$$

$$\frac{\partial L}{\partial w_1} = \frac{1}{N} \sum ((w_0 + w_1 x_n - t_n) \times x_n)$$

$$\text{temp}_0 = w_0^{old} - \alpha \frac{\partial L}{\partial w_0}$$

$$\text{temp}_1 = w_1^{old} - \alpha \frac{\partial L}{\partial w_1}$$

Size, # be, N floors | Age

$d = \#$ features

$x_m^i = i$ th feature x

$$x_2 = \begin{bmatrix} 1416 \\ 2 \\ 2 \\ 40 \end{bmatrix}$$

$$x_2^3 = 2$$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2^1 & x_2^2 & & x_2^d \\ \vdots & & & & \\ 1 & x_N^1 & x_N^2 & \cdots & x_N^d \end{bmatrix}$$

$$f = w^T x$$

$$\sigma \cdot p \quad \text{mv} \quad V$$

Repeat

$$tmp_j = w_j^{old} - \alpha \frac{\partial L}{\partial w_j}$$

$$w_j^{new} = tmp_j$$

until convergence

$$L = \sum (f(x_i) - t_n)^2$$

$$\text{temp}j = w_j^{\text{old}} - \alpha \sum \left(f(x_n : w) - t_n \right) x_n^j)$$