$$\boxed{\begin{aligned}&\frac{\partial \sum e_i^2}{\partial \theta_j} = 2\sum_{i=1}^{N}\Big(y_i - \underbrace{(x_i^0\theta_0 + \dots x_i^d\theta_d)}_{\hat{y}_i}\Big)\left(-x_i^j\right)\\ &\text{For } j=0,\quad x_i^{j=0} = 1 \ \forall\, i \in (1,\dots N)\end{aligned}}$$

$\sum_{i=1}^{N} e_i^2 \uparrow$ as # examples increases $\Rightarrow$ Residual Sum of Squares (RSS) or Sum of Squared Errors (SSE) increase

Sol^n: Use Mean squared error i.e. $MSE = \frac{RSS \text{ or } SSE}{N \ (\#\text{examples})}$

$$\Rightarrow \frac{\partial MSE(\theta_1, \dots \theta_d)}{\partial \theta_j} = \frac{2}{N}\sum_{i=1}^{N}\left(y_i - (x_i^0\theta_0 + \dots x_i^d\theta_d)\right)\left(-x_i^j\right)$$

G.D. worked out example

x	y
1	1
2	2
3	3

$$\hat{y} = \theta_0 + \theta_1 x$$

$$e_i = y_i - \hat{y}_i = y_i - (\theta_0 + \theta_1 x_i)$$

$$e_1 = 1 - \theta_0 - \theta_1 \qquad e_1^2 = (1 - \theta_0 - \theta_1)^2$$

$$e_2 = 2 - \theta_0 - 2\theta_1 \qquad \vdots$$

$$e_3 = 3 - \theta_0 - 3\theta_1$$

$$\vdots$$

$$\sum e_i^2 = f(\theta_0, \theta_1) = (1 + \theta_0^2 + \theta_1^2 - 2\theta_0 - 2\theta_1 + 2\theta_0\theta_1) + (4 + \theta_0^2 + 4\theta_1^2 - 4\theta_0 + 4\theta_0\theta_1 - 8\theta_1)$$

$$+ (9 + \theta_0^2 + 9\theta_1^2 - 6\theta_0 - 18\theta_1 + 6\theta_0\theta_1)$$

$$= 14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1$$

$$\boxed{\frac{\partial}{\partial\theta_0}\left(\frac{\sum e_i^2}{N}\right) = \frac{2}{N}\sum \left(y_i - (\theta_0+\theta_1 x_i)\right)(-1)}$$

&

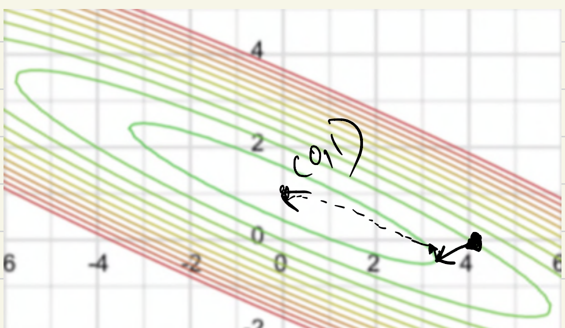
$$\boxed{\frac{\partial}{\partial\theta_1}\left(\frac{\sum e_i^2}{N}\right) = 2\sum \left(y_i - (\theta_0+\theta_1 x_i)\right)(-x_i)}$$

1. Randomly init θ_0, θ_1 to $(4, 0)$
2. $\alpha = 0.1$
3. Till convergence.

Iteratⁿ 1

$$\theta_0 = \theta_0 - \alpha \frac{2}{N} \sum \overbrace{\left(y_i - (\theta_0+\theta_1 x_i)\right)}^{e_i}(-1) = 4 - \frac{.2}{3}\left\{(1-(4+0))(-1) + (2-(4+0))(-1) + (3-(4+0))(-1)\right\}$$

$$= 4 - \frac{.2}{3}\left\{3+2+1\right\} = 4 - .2 \times 2 = 3.6$$



$$\theta_1 = \theta_1 - \frac{2\alpha}{N}\left[\sum (y_i - (\theta_0 + \theta_1 x_i))(-x_i)\right]$$

$$= 0 - \frac{.2}{3}\left[(1-(4+0))(-1) + (2-(4+0))(-2) + (3-(4+0))(-3)\right]$$

$$= 0 - \frac{.2}{3}\left[3 + 4 + 3\right] = -.67$$

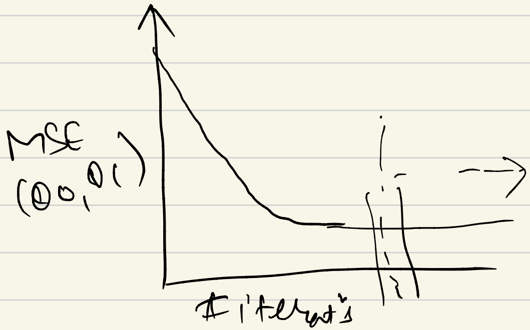
Iterat^n 2

$$\theta_0 = 3.6 - \frac{.2}{3}\left\{(1-(3.6-.67\times1))(-1) + (2-(3.6-.67\times2))(-2) + (3-(3.6-.67\times3))(-3)\right\}$$

$$= 3.54$$

$$\theta_1 = -.55$$

Iterat^n 500 $\theta_0 \simeq 1$; $\theta_1 \simeq 0$



Stopping criterion

when MSE at iterat^n N and N+1 are very close.

S.G.D. worked out example

x	y
1	1
2	2
3	3

$$\hat{y} = \theta_0 + \theta_1 x$$

$$e_i = y_i - \hat{y}_i = y_i - (\theta_0 + \theta_1 x_i)$$

$$e_1 = 1 - \theta_0 - \theta_1 \qquad e_1^2 = (1 - \theta_0 - \theta_1)^2$$

$$e_2 = 2 - \theta_0 - 2\theta_1 \qquad \vdots$$

$$e_3 = 3 - \theta_0 - 3\theta_1$$

$$\vdots$$

$$\Sigma e_1^2 = f(\theta_0, \theta_1) = (1 + \theta_0^2 + \theta_1^2 - 2\theta_0 - 2\theta_1 + 2\theta_0\theta_1) + (4 + \theta_0^2 + 4\theta_1^2 - 4\theta_0 + 4\theta_0\theta_1 - 8\theta_1)$$

$$+ (9 + \theta_0^2 + 9\theta_1^2 - 6\theta_0 - 18\theta_1 + 6\theta_0\theta_1)$$

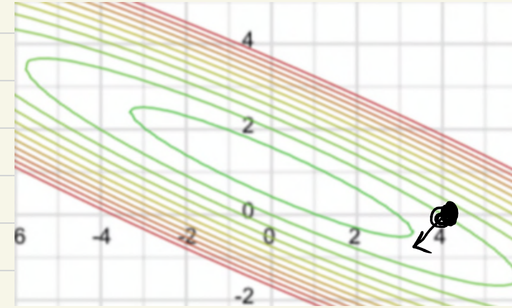
$$= 14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1$$

Randomly Initialize $(\theta_0, \theta_1) = (4, 0)$

Shuffle data

x	y
2	2
3	3
...	...

$\alpha = .1$



Update 1 : see first example x y

2 2

$$L_1^2 = (2 - \theta_0 - 2\theta_1)^2$$

For 1th example

$$\theta_0 = \theta_0 - \alpha \{ 2 (y_i - (\theta_0 + \theta_1 x_i)) (-1) \} = 4 - \alpha \{ 2 (2-4)(-1) \} = 4 - \alpha \times 4$$

$$= 4 - 4\alpha = 3.6$$

$$\theta_1 = \theta_1 - 2\alpha \{ (2 - (4+0)) (-2) \} = 0 - .2 \{4\} = -0.8$$

For 2nd example;

$$\theta_0 = 3.6 - .2\left\{\left(3 - (3.6 - 0.8 \times 3)\right)(-1)\right\} = 3.6 - .2\left\{(3 - 1.2)(-1)\right\}$$

$$= 3.6 + .2 \times 1.8$$

$$= 3.6 + .36 = 3.96$$

$$\theta_1 = .28$$

For 3rd example

$\theta_0 = \cdots = 3.3$, $\theta_1 = -.36$

Now, 3 iterat^{n}s and $\underline{\underline{1\ \text{epoch}}}$ is complete

updates to parameters

seen all data