Gradients

We wish to minimise $\frac{1}{2}SSE$ (our definition of SSE for mathematical convenience)

$$SSE = \frac{1}{2}(Y - YP)^2$$

Gradient of intercept

For finding gradient of $\boldsymbol{a}$ towards SSE, differentiate SSE with respect to **a**

	A	B	C	D	E	F	G
1		a = **0.45**		b= **0.75**		YP = a+Bx	
2		*Sq Ft*	*Price$*	*X*	*Y*	*YP*	*(1/2)SSE*
3		1100	199000	0.00	0.00	0.45	0.10125
4		1400	245000	0.22	0.22	0.62	0.077368
5		1425	319000	0.24	0.58	0.63	0.001154
6		1550	240000	0.33	0.20	0.70	0.125486
7		1600	312000	0.37	0.55	0.73	0.016062
8		1700	279000	0.44	0.39	0.78	0.078006
9		1700	310000	0.44	0.54	0.78	0.02989
10		1875	308000	0.57	0.53	0.88	0.061751
11		2350	405000	0.93	1.00	1.14	0.010432
12		2450	324000	1.00	0.61	1.20	0.175945
13	MIN	1100	199000	0	0	**Total SSE=**	**0.677345**
14	MAX	2450	405000	1	1		
15	RANGE	1350	206000	1	1		

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = \frac{\partial}{\partial \boldsymbol{a}}\left[\frac{1}{2}(Y - YP)^2\right]$$

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = \left[\frac{1}{2} \times \frac{\partial}{\partial \boldsymbol{a}}(Y - YP)^2\right]$$

Diff X^2

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = \frac{1}{2} \times 2 \times (Y - YP) \times \frac{\partial}{\partial \boldsymbol{a}}(Y - \boldsymbol{YP})$$

[diff of x^2 = 2*x^(n-1)* diff of x wrt x

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = (Y - YP) \times \frac{\partial}{\partial \boldsymbol{a}}[Y - (a + bX)] \text{ .... replacing } \boldsymbol{YP} \text{ by } (a + bX)$$

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{a}}Y - \frac{\partial}{\partial \boldsymbol{a}}(a + bX)\right]$$

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{a}}Y - \frac{\partial}{\partial \boldsymbol{a}}a - \frac{\partial}{\partial \boldsymbol{a}}(bX)\right]$$

$$\frac{\partial}{\partial \boldsymbol{a}}SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{a}}Y - \frac{\partial}{\partial \boldsymbol{a}}a - X\frac{\partial}{\partial \boldsymbol{a}}b\right] \text{ .... assuming } X \text{ as constant}$$

$$\frac{\partial}{\partial a} SSE = (Y - YP) \times [0 - 1 - X \times 0]$$

$$\frac{\partial}{\partial a} SSE = (Y - YP) \times [0 - 1 - 0]$$

$$Gradient\ of\ \boldsymbol{a}\ towards\ SSE = \frac{\partial}{\partial \boldsymbol{a}} SSE = -(\boldsymbol{Y} - \boldsymbol{YP})$$

Gradient of regression coefficient

For finding gradient of ***b*** towards SSE, differentiate SSE with respect to **b**

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = \frac{\partial}{\partial \boldsymbol{b}} \left[\frac{1}{2}(Y - YP)^2\right]$$

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = \left[\frac{1}{2} \times \frac{\partial}{\partial \boldsymbol{b}}(Y - YP)^2\right]$$

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = \frac{1}{2} \times 2 \times (Y - YP) \times \frac{\partial}{\partial \boldsymbol{b}}(Y - \boldsymbol{YP})$$

$$\frac{\partial}{\partial b} SSE = (Y - YP) \times \frac{\partial}{\partial b}[Y - (a + bX)] \text{ .... replacing } \boldsymbol{YP} \text{ by } (a + bX)$$

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{b}} Y - \frac{\partial}{\partial \boldsymbol{b}}(a + bX)\right]$$

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{b}} Y - \frac{\partial}{\partial \boldsymbol{b}} a - \frac{\partial}{\partial \boldsymbol{b}}(bX)\right]$$

$$\frac{\partial}{\partial b} SSE = (Y - YP) \times \left[\frac{\partial}{\partial b} Y - \frac{\partial}{\partial b} a - X \frac{\partial}{\partial b} b\right] \text{ .... assuming } X \text{ as constant}$$

$$\frac{\partial}{\partial b} SSE = (Y - YP) \times [0 - 0 - X \times 1]$$

$$\frac{\partial}{\partial \boldsymbol{a}} SSE = (Y - YP) \times \left[\frac{\partial}{\partial \boldsymbol{a}} Y - \frac{\partial}{\partial \boldsymbol{a}} a - X \frac{\partial}{\partial \boldsymbol{a}} b\right] \text{.... assuming } X \text{as constant}$$

$$\frac{\partial}{\partial \boldsymbol{a}} SSE = (Y - YP) \times [0 - 1 - X \times 0]$$

$$\frac{\partial}{\partial \boldsymbol{b}} SSE = (Y - YP) \times [0 - 0 - \boldsymbol{X}]$$

$$Gradient\ of\ \boldsymbol{b}\ towards\ SSE = \frac{\partial}{\partial \boldsymbol{b}} SSE = -(\boldsymbol{Y} - \boldsymbol{YP})\boldsymbol{X}$$

Updation

$$Updated\ weight\ a_{n+1} = old\ weight\ (a_n) - learning\ rate \times \frac{\partial}{\partial \boldsymbol{a}} SSE$$

Update in **a**

$$Updated\ weight\ a_{n+1} = 0.45 - 0.01 \times 3.30 = \boldsymbol{0.417 = 0.42}$$

$$Updated\ weight\ b_{n+1} = old\ weight\ (b_n) - learning\ rate \times \frac{\partial}{\partial \boldsymbol{b}} SSE$$

Update in **b**

$$Updated\ weight\ b_1 = 0.75 - 0.01 \times 1.55 = \boldsymbol{0.7345 = 0.73}$$

Momentum (beta)

Updating a

$$V_{updated\ (Da),n+1} = beta \times V_{AVG\ OF\ PREVIOUS\ \boldsymbol{Da}'s\ till\ 'n-1'} + (1 - beta) \times D_{\boldsymbol{a}\ (n)}$$

$$Updated\ weight\ \boldsymbol{a}_{n+1} = \boldsymbol{a}_n - learning\ rate\ \times V_{updated(D_{\boldsymbol{a}}),n+1}$$

Updating b

$$V_{updated\ (Db),n+1} = beta \times V_{AVG\ OF\ PREVIOUS\ \boldsymbol{Db}'s\ till\ 'n-1'} + (1 - beta) \times D_{\boldsymbol{b}\ (n)}$$

$$Updated\ weight\ \boldsymbol{b}_{n+1} = \boldsymbol{b}_n - learning\ rate\ \times V_{updated(D_{\boldsymbol{b}}),n+1}$$