**Assignment 6**

Referral id: SIRSS1088

NAME: ONASVEE BANARSE

EMAIL : 2obanarse@gmail.com

COLLEGE: AISSMS IOIT

GitHub : <https://github.com/ORION-22/RegexSoftware_ASSIGNMENT.git>

**Q1. Calculate/ derive the gradients used to update the parameters in cost function optimization for simple linear regression.**

⇒

Gradient Descent:

Gradient descent is an iterative optimization algorithm to find the minimum of a function.

The equation for simple regression is

y= 𝑎1 \* 𝑥 + 𝑎0

we know that cost or error(e) = 𝑦 ^ − 𝑦

for n data points:

𝑓(𝑎) = 1 𝑛 ∑ (𝑦 ^ − 𝑦) 𝑛 2 𝑖=1

𝑓(𝑎) = 1 𝑛 ∑ (𝑦 ^ − (𝑎1 ∗ 𝑥 + 𝑎0)) 𝑛 2 𝑖=1

α= learning rate or the size of the step we take towards finding the optimal fit line

𝑑𝑓(𝑎)/𝑑𝑎0

𝑝𝑎𝑟𝑡𝑖𝑎𝑙 𝑑𝑒𝑟𝑖𝑣𝑎𝑡𝑖𝑣𝑒 𝑜𝑓 𝑓(𝑎) 𝑤. 𝑟.𝑡 𝑎0 𝑤𝑖𝑙𝑙 𝑔𝑖𝑣𝑒 𝑡ℎ𝑒 𝑣𝑎𝑙𝑢𝑒 𝑜𝑓 𝑝𝑎𝑟𝑎𝑚𝑒𝑡𝑒𝑟 𝑎0

𝑎0= 2 𝑛 ∑ (𝑦 ^ − (𝑎1 ∗ 𝑥 + 𝑎0)) 𝑛 𝑖=1

𝑑𝑓(𝑎)/ 𝑑𝑎1 𝑝𝑎𝑟𝑡𝑖𝑎𝑙 𝑑𝑒𝑟𝑖𝑣𝑎𝑡𝑖𝑣𝑒 𝑜𝑓 𝑓(𝑎) 𝑤. 𝑟.𝑡 𝑎1 𝑤𝑖𝑙𝑙 𝑔𝑖𝑣𝑒 𝑡ℎ𝑒 𝑣𝑎𝑙𝑢𝑒 𝑜𝑓 𝑝𝑎𝑟𝑎𝑚𝑒𝑡𝑒𝑟

𝑎1 𝑎1= 2 𝑛 ∑ 𝑥(𝑦 ^ − (𝑎1 ∗ 𝑥 + 𝑎0)) 𝑛 𝑖=1

New 𝑎0= 𝑎0 − 𝑎0 ∗ α

New 𝑎1= 𝑎1 − 𝑎1 ∗ α

**Q2. What does the sign of gradient say about the relationship between the parameters and cost function?**

⇒

The cost function is a function of the parameters and when the sign is positive then the step will decrease as seen below:

New 𝑎0= 𝑎0 − [+𝑣𝑒 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α

when the sign is negative then the step will increase as seen below:

New 𝑎0= 𝑎0 − [−𝑣𝑒 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α

New 𝑎0= 𝑎0 + [ 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α

**Q3. Why Mean squared error is taken as the cost function for regression problems.**

⇒

MSE or Mean Squared Error is used to check how close predictions made by the model are to actual values. It calculates the error as actual - prediction and squares the difference to eliminate the negative values.

The lower the MSE, the closer is prediction to actual. In Regression models, a lower MSE usually indicates a better fit.

**Q4. What is the effect of learning rate on optimization, discuss all the cases?**

⇒

In an ideal scenario with an optimal learning rate, the cost function value will be minimized rather quickly.

If we take a large learning rate then the cost function value will be minimized very quickly but will settle at a value that is not the lowest.

If we take a lower than optimal learning rate, then even after substantial iterations the cost function will not minimize sufficiently and will take longer time.