CO544 - Machine Learning & Data Mining
Tutorial 2
E/16 319 Rathnayake R.P.V.N

Q1.
Prediction can be taken as 2.5 because there are more points at 2.5 marks. Because It shows some good precision around 2.5. But give GPA as 2.35 to 2.75 range because it has this range. But by considering precision 2.5 can be taken

Q2.
Third function which gets the sum of squares of the difference between actual and predicted would be a good candidate function. In First function it takes only the difference. That will subtract the minus difference from the value and give misleading information. In the second function, although it takes the absolute value, It have same weight for higher and lower differences. But in Third one it takes the square of difference and gives high weight to higher differences than smaller ones and also does not deduct the minus differences.

Q3
1)Why does this move in the direction of steepest descent?
Because in the gradient descent equation the $\theta 1$ changes in the direction of steepest direction

2)What are potential problems with having a too big learning rate (α) ?
Having larger learning rate may guide to fast convergence by it may cause overshoot and causes dramatic updates and go into divergence too.

3)What are potential problems with having a too small learning rate?
Having a small learning rate will take a very long time to reach the minimum point.

4)What would we do if we wanted to maximize J(θ instead?
Change the gradient equation by change the subtraction to addition in the equation

Relative questions

1. 1 hour
2. Last part of Q3 is bit confusing, What would we do if we wanted to maximize J(θ instead?