# Assignment 2 Question 3

Sheen Thusoo

## Part a)

The overall purpose of Gradient Descent is to minimize an objective function in an iterative or repetitive manner; essentially, it is an optimization algorithm for finding where the lowest point of a function occurs. The gradient of a function is the slope of that function at a specific point. In this case, we can imagine our function to be in the shape of a hill with many bumps. We start at the beginning of the hill and want to move in the direction that will get us to a lower point. We do this by calculating the gradient or slope at a point close-by on the hill and moving in the direction of steepest descent (downhill). We also need to decide how big of a step we should take in this direction - this is done by choosing a step size, $\lambda$. This step size can be chosen in many ways: it can be fixed, it can be a sequence, or we can find an optimal value for it. Ideally, we want the step size to be larger at first, since we are further away from the lowest point, and then as we approach this lowest point we want it to become smaller so that we do not "jump over" it. We need to choose a step size in which we can get to the lowest point quickly, but also not overshoot it. If the step size is too small, it will take us a very long time to get to the lowest point on the hill. Once we have our gradient, the direction, and the step-size, we move towards a lower value in the function. We then make note of which point we are on the hill. This process is repeated until we reach the minimum point. We figure out that we have reached a minimum when we cannot find a slope that takes us any lower, we can only go higher. At the end, we return the point which yields the lowest part of the hill. An important thing to define is where we want to start on the hill as this can affect where the algorithm takes us. Suppose there are many low points on the hill, but only one where we are the lowest. If we start in a different area, the algorithm could take us to a point which is a minimum in a local area but is not the minimum point of the whole hill. When we define this algorithm, we can also set a maximum number of iterations or repetitions that we want to do before giving up. Sometimes, the absolute minimum of the function cannot be reached if the algorithm reaches its maximum iteration number. We set this number because we do not want our algorithm to go on for a very long time. All in all, this algorithm finds the gradient or slope at a point on the function and moves in a downwards direction until we reach a new point where we find the slope again and so on. We keep doing this until we reach the lowest point.