Homework2 For Machine Learning

Vergil/Zijun Li李子骏

1. **Suppose you are running gradient descent. As we reach the local minima, the gradient descent needs to take smaller steps to avoid overshooting to the other side. Do we need to reduce the learning rate α to achieve that? If yes, please explain. If not, please clarify how we can achieve taking smaller steps. (2 points).**

   Yes.

   By reducing the learning rate α, we can get smaller steps as it controls the size of the steps taken in each iteration.

2. **Calculate the partial derivative of the following expression with respect to x: $x^3 + 3xy + y^2 + 1$ (2 points).**

$$\frac{\partial}{\partial x}(x^3 + 3xy + y^2 + 1) = 3x^2 + 3y$$

3. **Normal equation method requires computing $X, X^T, (X^T X)^{(-1)}$ and y to be able to compute θ. Suppose number of training examples m = 8 and the number of features is 4 (excluding the all 1 feature). Write down the dimension of each of the following 5 matrices: $X, X^T, (X^T X)^{(-1)}$, y,and θ.(4 points)**

   X: 8x5 matrix(8 examples * 4+1 features)

   X^T: 5x8 matrix(4+1 features * 8 examples)

   $(X^T X)^{-1}$: 5x5 matrix(4+1 features)

   y: 8x1 matrix(8 examples)

   θ: 5x1 matrix(4+1 features)

4. **You ran gradient descent for 32 iterations with learning rate value of 0.1. However it seems that the cost function initially decreased but then it's value started increasing with iterations. What would you do in this case?** B

   (A) Just wait for more iterations, the problem should fix itself automatically

   (B) Reduce the learning rate and restart

   (C) Increase the learning rate and restart

   (D) Gradient descent does not seem suitable here. Go to the normal equation method. (3 points)

5. **Why do we use feature scaling:**D

   (A) Gradient descent can produce an incorrect answer without proper feature scaling

   (B) Gradient descent can go in an infinite loop without proper feature scaling

(C) Gradient descent can get stuck in local minima without proper feature scaling

(D) [✓ Gradient descent can take a long time to finish without proper feature scaling]

6. **The price of a diamond goes up as the carat weight goes up. Following is a sample table given to you:**

Carat Weight	Price in thousands
1	1
2	4
3	8
4	15
8	62
14	193

Your goal is to build a model for predicting the diamond price given it's carat weight. What would be your hypothesis?

I'd like to use linear regression here as we only have one feature(weight) here.

$$h_\theta(x) = \theta_0 + \theta_1 * x$$

- $h_\theta(x)$ is the predicting the diamond price
- θ_0 and θ_1 are the parameters of linear regression
- x is the Carat Weight