DS 100: Principles and Techniques of Data Science	Date: March 2, 2018

Discussion #6

Name:

Loss Functions

1. Recall the loss functions discussed during lecture.Discuss the advantages and drawbacks of each of the following loss functions:

   a Squared loss: $L(\theta, y) = (y - \theta)^2$

   b Absolute Loss: $L(\theta, y) = |y - \theta|$

   c Huber Loss:
   $L(\theta, y) = \begin{cases} (y-\theta)^2 & |y-\theta| < \alpha \\ \alpha(|y-\theta| - \alpha/2) & \text{otherwise} \end{cases}$

Loss Minimization

Consider the following loss function.

$$L(\theta, x) = \begin{cases} 4(\theta - x) & \theta \geq x \\ x - \theta & \theta < x \end{cases}$$

2. Draw out this loss function about a point x. Is it convex?

3. Given a sample of $x_1, ..., x_n$, find the optimal θ^* that minimizes the the average loss.

Gradient Descent

4. Given the following loss function and x, y, θ^t, write out the update equation for θ^{t+1}.

$$L(\theta, x, y) = \sum_{i=1}^{n} \theta^2 * x_i^2 - log(y_i)$$

Modeling

5. We wish to model exam grades for DS100 students. We collect various information about student habits, such as how many hours they studied, how many hours they slept before the exam, how many lectures they attended and observe how well they did on the exam. Propose a model to predict exam grades and a loss function to measure the performance of your model.

6. Suppose we collected even more information about each student, such as their eye color, height, and favorite food. Do you think adding these variables as features would improve our model?

Convexity

7. Convexity is a very important and has many useful properties that apply to many relevant areas, including machine learning. It allows optimization problems to be solved more efficiently and for global optimums to be realized. This question will explore the notion of convexity. There are three ways to define convexity.

   a For all $\lambda \in [0, 1], x, y,$ we have $f(\lambda x + (1 - \lambda)y \leq \lambda f(x) + (1 - \lambda)f(y)$

   b For all x, y we have $f(y) >= f(x) + f'(x)(y - x)$

   c For all x, we have $f''(x) >= 0$

   Describe the geometric interpretation/meaning of each of the definitions. What assumptions do each of the definitions above make on the function?

8. Find a counterexample for the claim that the composition of two convex functions is also convex. $h = g(f(x))$

9. Prove that the composition of a convex function f and a convex non decreasing function g is also convex. $h = g(f(x))$
   Hint: Show the following is true

$$(g \circ f)(\lambda x + (1-\lambda)y) \leq \lambda(g \circ f)(x) + (1-\lambda)(g \circ f)(y)$$

10. Show that the union of two convex sets is not necessarily convex.