# CSE 151A - Homework 05

Due: Wednesday, May 6, 2020

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope on Wednesday at 11:59 p.m.

**Essential Problem 1.**

Consider the function $h(x) = (x^3 - x + 1)^2$. It is easy to verify that this is a non-convex function. We know that for a convex function with the appropriate choice of the learning rate, gradient descent converges to the global minimum. However for a non-convex function, since there can be multiple local minima, gradient descent can converge to any local minima. An intuitive workaround for this problem then is to perform gradient descent from multiple initial values of $x$ and then find the minimum among all the local minima in order to get a better estimate of the global minima.
Run 5 iterations of gradient descent with learning rate 0.05 for $h(x)$ with the following two initial values of $x$ and find the global minima of $h(x)$ (given that $h(x)$ has only two local minima):

1. $x_0 = 0$
2. $x_0 = -1$

**Note**: you should run at least the first two iterations of gradient descent "by hand", without using code, so as to get a feeling for the procedure. You may use code (including the code in the demo notebook posted during lecture) to do the remaining iterations.

**Essential Problem 2.**

Suppose you have gathered the data below:

| Feature 1 | Feature 2 | Class |
|---|---|---|
| -1 | 2 | 1 |
| 2 | -3 | 1 |
| 1 | 0 | 1 |
| -3 | 1 | -1 |
| 1 | 1 | -1 |

This data is just randomly generated and doesn't mean anything, but you can think of each row as being a person, and each feature is a different measurement about that person (like their height or weight).

Suppose you use a logistic regression model to predict the class label. You train your model using gradient descent to find the weight vector $\vec{w}$.

Below are two vectors, $\vec{w}^{(1)}$ and $\vec{w}^{(2)}$. One of them is the weight vector found by gradient descent, and the other is a vector I made up. Which is the vector found by gradient descent? Explain how you know this.

$$\vec{w}^{(1)} = (-0.31, -1.29, 3.90)^T$$
$$\vec{w}^{(2)} = (-0.22, -1.41, 3.95)^T$$

**Essential Problem 3.**

For each of the following functions, determine if it is convex or not using the second derivative test. Show your work in each case.

**a)** $f(x) = 3x^3 + 2x - 4$

**b)** $f(x) = \dfrac{e^x + e^{-x}}{2}$

**c)** $f(x, y) = x^2 + y^2 - 5xy + 10x + 12y - 42$

**d)** $f(\vec{x}) = \|\vec{x}\|^2$

**Plus Problem 1.** (6 plus points)

The gradient descent update rule for minimizing a function $R(h)$ is:

$$h_{\text{next}} = h_{\text{prev}} - \alpha \frac{dR}{dh}(h_{\text{prev}}).$$

We said in class that the sign of $dR/dh$ is meaningful: if it is positive we should move to the left, and if it negative we should move to the right.

Why is the *magnitude* of the derivative useful, too? That is, what is wrong with using the update rule:

$$h_{\text{next}} = h_{\text{prev}} - \alpha \cdot \operatorname{sign}\left(\frac{dR}{dh}(h_{\text{prev}})\right),$$

where $\operatorname{sign}(\cdot)$ returns the sign of its argument as either zero or one. For instance, $\operatorname{sign}(-4) = -1$ and $\operatorname{sign}(42) = 1$.

*Note: I'll offer a larger-than-average number of plus points in next week's homework.*