Introduction to Machine Learning
Homework 0: Intuition

Prof. Linda Sellie

In machine learning, we are looking to find the "best" function that allows us to predict new examples. We typically define the "best" function as one that maximizes or minimizes an objective function. In this homework, I would like you to start developing an intuitive idea of how a function can be maximized or minimized. In class, we will spend time thinking about what objective function "makes sense" for our problem, and then we will discover ways to maximize or minimize the function.

1. If $f(x) = 7x^2 + 8x + 10$, then $f(10) = 790$.

   The best straight line approximation to $f(x)$ at any point x is the line $14x + 8$.

   Without evaluating the function at $f(10+0.01)$ and $f(10-0.01)$, can you determine if $f(10+0.01) < f(10)$ or is $f(10-0.01) < f(10)$?

   Why or why not?

2. Suppose you decided to spend the day mountain climbing.

   Having spent a fun day climbing, you notice it is time to get back to your car which is in the valley at the base of the mountain (since the sun has gone down and you can no longer see the beautiful views, ... or anything else). Unfortunately, you have lost yourself in the woods. Describe an algorithm that allows you to get back as quickly as possible, where you assume that the side of the mountain you are on has only one valley.

   I am just looking for a simple 1 or 2-sentence intuitive answer for this question. (Don't overthink this question.)

3. If you knew the part of the mountain you are walking on doesn't have only one valley - would this make a difference?

   What could go wrong? (Don't overthink this question.)

4. Suppose you traveled to Los Vegas and you were playing a coin game on the street. You are pretty sure the coin was not fair. You would like to go back and win your \$50 back. If in 50 coin flips, there were 30 heads (and 20 tails). What is the most

likely probability of the coin having heads? Let p_h be the probability of the coin being heads.

What p_h would maximize $(p_h)^{30}(1-p_h)^{20}$?[1]

Since we know that $p_h \neq 1$ and $p_h \neq 0$ since both heads and tails occurred, would finding the p_h that maximizes $(p_h)^{30}(1-p_h)^{20}$ be the same as finding the p_h that maximizes $\log\big(\,(p_h)^{30}(1-p_h)^{20}\,\big) = 30\log(p_h) + 20\log(1-p_h)$?

Why or why not?

[1] Hint: Use calculus. The function is strictly concave which means it has a unique maximum value. You can get full credit for this part of the question by showing you thought about the answer.