CSC321H5 Homework 1.

Deadline: Thursday, Jan. 16, by 9pm

Submission: You must submit your solutions as a PDF file through MarkUs. You can produce the file however you like (e.g. LaTeX, Microsoft Word, scanner), as long as it is readable.

Late Submission: Please see the syllabus for the late submission criteria. If you are enrolling in the course late, please email Lisa (course coordinator) at lczhang@cs.toronto.edu

Question 1. Machine Learning

Part (a) – 3 pts

Which of these problems would you use machine learning to solve? Explain why.

- Determining the solution to a sudoku puzzle
- Determining whether a photo contains a cat
- Determining when the next train will arrive at Clarkson station
- Determining whether a piece of text is written by a child or an adult
- Determining the number of correct answers in a scantron sheet
- Determining how a robot vacuum cleaning should move

Part (b) – 3 pts

For each of the problems from question 1, identify whether the problem is a classification problem, regression problem, or none of the above. Explain your choice.

Question 2. Visualizing the Loss Function

In lecture, we visualized the linear regression cost function in weight space and saw that the contours were ellipses. Let's work through a simple example of that. In particular, suppose we have a linear regression model with two weights and no bias term:

$$y = w_1 x_1 + w_2 x_2$$

and the usual loss function

Suppose we have a training set consisting of $N = 3$ examples:

- $\mathbf{x}^{(1)} = (2, 0), t^{(1)} = 1$
- $\mathbf{x}^{(2)} = (0, 1), t^{(2)} = 2$
- $\mathbf{x}^{(3)} = (0, 1), t^{(3)} = 0$

Let's sketch one of the contours.

Part (a) – 4 pts

Write the cost in the form $\mathcal{E} = c_1(w_1 - d_1)^2 + c_2(w_2 - d_2)^2 + \mathcal{E}_0$

Part (b) – 5 pts

Since $c_1, c_2 > 0$, this corresponds to an axis-aligned ellipse. Sketch the ellipse by hand for $\mathcal{E} = 1$. Label the center and radii of the ellipse. If you've forgotten how to plot axis-aligned ellipses, see Khan Academy[1].

[1] https://www.khanacademy.org/math/algebra-home/alg-conic-sections/alg-center-and-radii-of-an-ellipse/v/conic-sections-intro-to-ellipses

Question 3. Gradient Descent Preview

We will be discussing gradient descent in lecture 2. Before we do, we would like you to work through some of the math that we will be using for the lecture.

Consider the model $y = wx + b$ from lecture. We wish to find w and b to minimize

$$\mathcal{E}(w, b) = \frac{1}{2N} \sum_{i=1}^{N} ((wx^{(i)} + b) - t^{(i)})^2$$

over some labeled data $(x^{(1)}, t^{(1)})$, $(x^{(2)}, t^{(2)})$, $\ldots$ $(x^{(N)}, t^{(N)})$.

Part (a) – 3 pts

Derive the quantity $\frac{\partial \mathcal{E}}{\partial w}$.

Note: If you haven't taken a course in multivariable calculus, you might not have seen the symbol ∂. Think of ∂ as similar to d, but is signifying that all other quantities other than w are thought of as fixed. So, think about finding $\frac{d\mathcal{E}}{dw}$, while fixing all other quantities like b, N, $x^{(i)}$, and $t^{(i)}$.

Part (b) – 3 pts

Derive the quantity $\frac{\partial \mathcal{E}}{\partial b}$.

Question 4. An Alternative Loss Function

Suppose that instead of using the square loss $\mathcal{L}(y, t) = \frac{1}{2}(y - t)^2$ in lecture, we wish to use the **absolute loss** $\mathcal{L}(y, t) = |y - t|$

Part (a) – 1 pts

Derive the cost function $\mathcal{E}(y, t)$, which is the average value of $\mathcal{L}$ over some labeled data $(x^{(1)}, t^{(1)})$, $(x^{(2)}, t^{(2)})$, $\ldots$ $(x^{(N)}, t^{(N)})$.

Part (b) – 3 pts

Derive the quantity $\frac{\partial \mathcal{E}}{\partial w}$.

Update Jan 13 You may find this function useful:

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x >= 0 \\ -1, & \text{if } x < 0 \end{cases}$$

Part (c) – 3 pts

Derive the quantity $\frac{\partial \mathcal{E}}{\partial b}$.

Part (d) – 2 pts

Is it possible to solve for the optimal w by setting $\frac{\partial \mathcal{E}}{\partial w} = 0$? Why or why not?