CSE 151: Machine learning **Winter 2019**

Homework 6

Submission instructions:

- Please type up your solutions.
- If a problem asks for a numerical answer, you need only provide this answer. There is no need to show your work, unless you would like to.
- Upload the PDF file for your homework to `gradescope` by 6pm on Tuesday February 19.

Part A: Convexity and linear classification

1. For each of the following functions of one variable, say whether it is convex, concave, both, or neither.
   (a) $f(x) = x^2$
   (b) $f(x) = -x^2$
   (c) $f(x) = x^2 - 2x + 1$
   (d) $f(x) = x$
   (e) $f(x) = x^3$
   (f) $f(x) = x^4$
   (g) $f(x) = \ln x$
2. Show that the matrix $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is not positive semidefinite.
3. Show that the matrix $M = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ is positive semidefinite.
4. For a fixed set of vectors $v_1, \ldots, v_n \in \mathbb{R}^d$, let M be the $n \times n$ matrix of all pairwise dot products: that is, $M_{ij} = v_i \cdot v_j$. Show that M is positive semidefinite.
5. For some fixed vector $u \in \mathbb{R}^d$, define
$$F(x) = \|x - u\|^2.$$
Is $F(x)$ a convex function of x? Justify your answer.
6. For some fixed vector $u \in \mathbb{R}^d$, define the function $F : \mathbb{R}^d \to \mathbb{R}$ by $F(x) = e^{u \cdot x}$.
   (a) What is the Hessian $H(x)$?
   (b) If F a convex function of x? Justify your answer.

7. Let $p = (p_1, p_2, \ldots, p_m)$ be a probability distribution over m possible outcomes. The *entropy* of p is a measure of how much randomness there is in the outcome. It is defined as

$$F(p) = -\sum_{i=1}^{m} p_i \ln p_i,$$

where ln denotes natural logarithm. Is this a convex function, or a concave function, or neither? Justify your answer.

8. Draw the decision boundary in $\mathbb{R}^2$ that corresponds to the prediction rule $\text{sign}(2x_1 - x_2 - 6)$. Make sure to clearly indicate where this boundary intersects the axes. Show which side of the boundary is classified as positive and which side as negative.

9. A particular labeled data of n points is randomly permuted and then the Perceptron algorithm is run on it, repeatedly cycling through the points until convergence. It converges after making k updates. For each of the following statements, say whether it is **definitely true** or **possibly false**, and give a brief reason.

   (a) The data set is linearly separable.

   (b) If the process were repeated with a different random permutation, it would again converge.

   (c) If the process were repeated with a different random permutation, it would again converge after making k updates.

   (d) k is at most n.

10. The Perceptron algorithm is run on a data set, and converges after performing $p+q$ updates. Of these updates, p are on data points whose label is -1 and q are on data points whose label is $+1$. What is the final value of the parameter b?

Part B: Programming problem

Perceptron algorithm. In this problem, you will code up the Perceptron algorithm and use it to classify the Iris data set.

(a) Write code for two functions:

- The first function takes as input parameters w, b of a linear classifier as well as a data point x, and returns the label for that point: $\text{sign}(w \cdot x + b)$. The label is either $+1$ or -1.
- The second function takes as input an array of data points and an array of labels (where each label is $+1$ or -1), and runs the Perceptron algorithm to learn a linear classifier w, b. The algorithm should begin by randomly permuting the data points.

In your writeup, give the code for these two functions.

(b) Load in the Iris data set. You can do this by simply invoking:

```
from sklearn import datasets
iris = datasets.load_iris()
x = iris.data
y = iris.target
```

The data has four features and three labels. Restrict it to features 1 and 3 (the second and fourth columns, sepal width and petal width) and to labels 0,1. Recode label 0 as -1, since this is what the Perceptron algorithm is expecting.

(c) Now run the Perceptron algorithm on the data. In your writeup, show a plot with the data points (where the two labels have different colors) and the resulting decision boundary.

(d) Now modify your code from part (a) to count the *number of updates* made by the Perceptron algorithm while it is learning. Run the algorithm 20 times and keep track of the number of updates needed each time. In your writeup, include a histogram of these values.