ECE-UY 4563: Introduction to Machine Learning
Midterm 2, Fall 2021

Prof. Sundeep Rangan

- The exam is two parts:
  - Written part: This file, three questions.
  - Python part: See the `Midterm2_2021_Python.ipynb` file. This part also three questions.

  Both files, along with the data, can be found on Brightspace.
- Submit answers to all six questions on Gradescope. You have at least 24 hours.
- You may use any resources on the web, class notes, and homework solutions. However, you may not ask friends or other classmates for help.
- Best of luck.

1. *Nonlinear optimization.* Given data (x_i, y_i), $i = 1, \ldots, N$, you try to fit a model of the form:

$$\widehat{y}_i = \sum_{j=1}^{M} \frac{a_j}{1 + \exp(-b_j x_i)},$$

for parameters a_j and b_j, $j = 1, \ldots, M$. You wish to minimize the squared error:

$$J = \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2.$$

(a) Find the gradient components, $\partial J / \partial a_j$ and $\partial J / \partial b_j$.

(b) Complete the following python function

```
def Jeval(a,b,...):
    ...
    return J, Jgrada, Jgradb
```

that computes J and $\nabla_a J$ and $\nabla_b J$. You need to complete the arguments of the function. To receive full credit, avoid using for loops.

2. *Linear SVM.* A linear SVM is trained on a large number of points. Four of the points are as follows:

x_{i1}	0	2	2	4	$\cdots$
x_{i2}	0	0.5	2	3	$\cdots$
y_i	-1	-1	1	1	$\cdots$

After training an SVM on all the data points, you get the classifier:

$$\widehat{y} = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0, \end{cases} \qquad z = 0.75x_1 + x_2 - 3.$$

(a) Draw a scatter plot of the four data points using different markers for the two classes. Also draw the classifier boundary where $z = 0$.

(b) Compute the hinge loss, $\epsilon_i = \max\{0, 1 - z_i y_i\}$ for each point.

(c) What is the minimum distance of the four samples to the classifier boundary line $z = 0$?

(d) Suppose the data sample at $(0, 0)$ was shifted to $(0.1, 0.1)$ and the SVM were retrained with all the other training samples remaining the same. Would the classifier shift boundary shift to the left, right, or stay the same? Explain.

3. *Neural Networks:* Consider a neural network with:

   - A scalar input x
   - A single hidden layer with weights W_j^{H}, biases b_j^{H}, and ReLU activations
   - An output layer with a scalar output $\widehat{y}$, weights W_j^{O}, bias b^{O} and a linear activation.

   The network defines a mapping $\widehat{y} = f(x, \theta)$ where θ is the set of parameters.

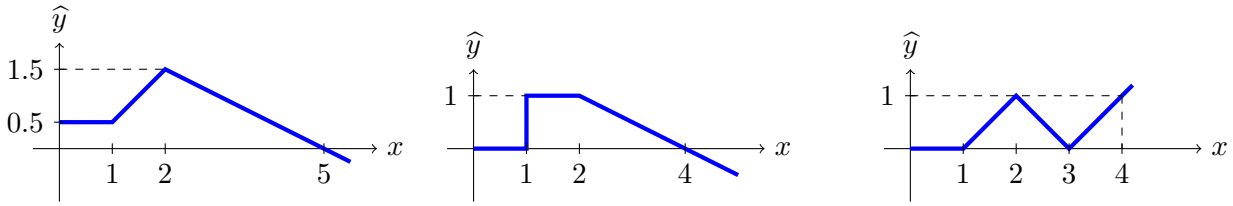
   (a) Write the equations for the neural network mapping x to $\widehat{y}$ in terms of the weights and biases.

   (b) Suppose that the neural network has $M = 2$ hidden units with hidden weights:

   $$W_j^{\mathrm{H}} = 1, \quad j = 1, \ldots, M.$$

   Fig. 1 below shows three functions, $\widehat{y} = f_i(x)$, $i = 1, 2, 3$. For each $f_i(x)$:

   - State if $f_i(x)$ can be the output of the neural network.
   - If not, explain why. One sentence is enough.
   - If so, find values for the parameters θ such that $f_i(x) = f(x, \theta)$.



(a) $f_1(x)$ (b) $f_2(x)$ (c) $f_3(x)$

Figure 1: Functions $\widehat{y} = f_i(x)$.