ECE-GY 6143: Introduction to Machine Learning
Midterm , Spring 2021

Name: **ID:**

Answer ALL questions. Exam is closed book. No electronic aids. However, you are permitted two cheat sheets, two sides each sheet. Any content on the cheat sheet is permitted. Part marks are given. If you do not remember a particular python command or its syntax, use pseudo-code and state what syntax you are assuming.

Please use a separate page for each question. Clearly write the question numbers on top of the pages. When you submit, please order your pages by the question numbers.

Each student is required to open Zoom and turn on their video, making sure the camera captures your hands and your computer screen/keyboard. The whole exam will be recorded. To clearly capture the video of exam taking, it is recommended that: A student can use an external webcam connected to the computer, or use another device (smartphone/tablet/laptop with power plugged in). Adjust the position of the camera so that it clearly captures the keyboard, screen and both hands. During the exam, you should keep your video on all the time. If there is anything wrong with your Zoom connection, please reconnect ASAP. If you cannot, please email ASAP.

During the exam, if you need to use the restroom, please send a message using the chat function in Zoom, so we know you left.

At 1:30PM, submit a single PDF to newclasses.nyu.edu, under the assignment named "Midterm Exam". You can use Adobe Scan APP/iPhone Notes APP/etc to scan your answers as a PDF file. If you have never used your phone to scan a document, give it a try now.

DEADLINE for submission is in 1:45PM. Before you leave the exam, it's your responsibility to make sure all your answers are uploaded.If you have technical difficulty and cannot upload till 1:40PM, email a copy to your proctor and the professor.

Pei Liu: peiliu@nyu.edu

Ashutosh Srivastava: as12738@nyu.edu

Vishaal Ranjan: vr1089@nyu.edu

Rahul Petkar: rsp431@nyu.edu

Best of luck!

1. (20 points) *linear regression.* Consider the following model.

   (a) (10 points) In a linear regression problem, you have trained a system to approximate a mapping from scaler x to scaler y as $\hat{y} = 2x + 1$ What is the residual sum of squares (RSS) over the following given test set? Show your work.

$x_1 = 0$	$y_1 = 2$
$x_2 = 3$	$y_2 = 5$
$x_3 = -7$	$y_3 = -15$
$x_4 = -2$	$y_4 = -3$

   (b) (10 points) You want to use a linear model to predict a vector $\mathbf{y} = (y_1, y_2)$ from a vector $\mathbf{x} = (x_1, x_2)$. Please derive the least square solution, where $RSS = \sum_{i=1}^{N} ||\mathbf{y}_i - \hat{\mathbf{y}}_i||^2$.

2. (15 points) *Least squares with a transformation.* Consider the following model for a scalar target $\hat{y}$ from features $\mathbf{x} = (x_1, x_2)$,

$$\hat{y} = f(x, \theta) = \begin{cases} \theta_0 + \theta_1(1 + \theta_2 x_2) & \text{if } x_1 < 0 \\ \theta_0 + \frac{\theta_1(1+\theta_2 x_2)}{1+x_1} & \text{if } x_1 \geq 0 \end{cases}$$

where $\theta = (\theta_0, \theta_1, \theta_2)$ are parameters.

(a) (10 points) Write this as a linear model. That is, find basis functions $\phi_i(\mathbf{x})$ and parameters β_i such that

$$\hat{y} = \sum_{i=1}^{p} \beta_i \phi_i(\mathbf{x}).$$

Use a minimum number p of basis functions. Write the linear model parameters β_i in terms of θ_j.

(b) (5 points) Given a linear parameter estimate $\boldsymbol{\beta}$, how do you find $\boldsymbol{\theta}$ in the original model?

3. (20 points) *Model order selection.* Suppose that a true function is

$$y = f_0(x) := \max\{x, 6\}.$$

(a) (10 points) Show that the true function is in the model class $f(x, \beta) = \beta_0 + \beta_1 x + \beta_2 \min(x, 6)$.

(b) (10 points) If we get training data (x_i, y_i), $i = 1, \ldots, n$ with $y_i = f_0(x_i)$, and fit a model as follows,

$$\hat{y} = f(x, \hat{\beta}) = \hat{\beta} x, \quad \hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}.$$

When the training data points is $\mathbf{x} = \{-1, 3, 12\}$, find the bias of the testing data point $x_{\text{ts}} = 2$, which is defined as $\text{Bias}(x_{\text{ts}}) = f(x_{\text{ts}}, \hat{\beta}) - f_0(x_{\text{ts}})$.

4. (10 points) *Maximum Likelihood Estimates(MLE).* Suppose $x_1, x_2, \ldots, x_n$ are i.i.d. samples from a uniform distribution $U(a, b)$, where $a < b$. I.e.,

$$p(x) = \frac{1}{b-a} \quad if \quad b \geq x \geq a; \quad p(x) = 0 \quad otherwise.$$

Find the MLE for parameter a and b.

5. (15 points) *Linear Classification.* You are given the following five training data points, (x_i, y_i), with binary class labels $y_i \in \{0, 1\}$:

x_i	0	1	2	3	4
y_i	1	1	0	0	1

Consider a classifier of the form,

$$\widehat{y}_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i < 0, \end{cases} \tag{1}$$

where z_i is some function of x_i.

(a) (10 points) Find β_0, β_1 such that with $z_i = \beta_0 + \beta_1 x_i$, the classifier (1) makes a minimum number of errors. Indicate which training data points, if any, are misclassified.

(b) (5 points) Find $\beta_0, \beta_1, \beta_2$ such that when $z_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$, the classifier (1) makes a minimum number of errors. Indicate which training data points, if any, are misclassified.

6. (20 points) *Gradients descent optimization.* Solve the following problems.

   (a) (10 points) Consider the function,

   $$J = \sum_{i=1}^{N} \ln(1 + e^{z_i}) - z_i y_i, \quad z_i = \sum_{j=1}^{p} \frac{1}{a_j + (x_i - b_j)^2}.$$

   Compute the gradient components, $\partial J / \partial a_j$ and $\partial J / \partial b_j$.

   (b) (10 points) You are given a python function $Jeval(a, b)$ that returns function J and the gradients, write a python gradient descent optimizer function with an adaptive step-size using the Armijo rule.

```
def grad_opt_armijo(Jeval, ainit, binit, nit, lr_init, lr_min):
    #ainit, binit are the initial value for a and b
    #nit is number of interations, lr stands for learning rate
    ...
return a_opt, b_opt, J_opt
```