CSCI-UA 9473 - Introduction to Machine Learning
Midterm I

Augustin Cosse

June 2022

Total: 27 points
Total time: 2h
General instructions: The exam consists of 2 questions (each question consisting itself of several subquestions). Once you are done, make sure to write your name on each page, then take a picture of all your answers and send it by email to acosse@nyu.edu. In case you have any question, you can ask those through the chat. Answer as many questions as you can starting with those you feel more confident with.

Question 1 (13pts)

1. [*5pts*] *Indicate whether the following statements are true or false*

   Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n} \left(t^{(i)} - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq s$$

   True / False   *As we increase s from 0, the training RSS will increase initially, and then eventually start decreasing in an inverted U-shape*

   True / False   *As we increase s from 0, the training RSS will decrease initially, and then eventually start increasing in an inverted U-shape*

   True / False   *As we increase s from 0, the training RSS will steadily increase*

   True / False   *As we increase s from 0, the training RSS will steadily decrease*

   Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n} \left(t^{(i)} - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

   for a particular value of λ

   True / False   *As we increase λ from 0, the variance will increase initially, and then eventually start decreasing in an inverted U-shape*

   True / False   *As we increase λ from 0, the variance will decrease initially, and then eventually start increasing in an inverted U-shape*

   True / False   *As we increase λ from 0, the variance will steadily increase*

   True / False   *As we increase λ from 0, the variance will steadily decrease*

2. *Explain how the binary classifier can be extended to a multiclass classification problem. Give three possible extensions and provide the associated pseudo code [5pts]*

3. *Explain how K-fold cross validation is implemented [3pts]*

Question 2 (14pts)

1. *Indicate whether the following statements are true or false [5pts]*

   True / False   Increasing the number of hidden layers in a neural network will decrease the bias

   True / False   Increasing the number of hidden layers in a neural network will increase the bias

   True / False   Using gradient descent to train a neural network will always return the global minimum of the loss

   True / False   When training a neural network through batch gradient descent, the value of the loss will always decrease

   True / False   When training a neural network through stochastic gradient descent, the value of the loss will always decrease

   True / False   In convolutional neural networks, pooling layers are used to reduce the dimension of the feature map

2. *[4pts] Let $\sigma(\boldsymbol{\beta}^T \tilde{\boldsymbol{x}})$ denote the two dimensional logistic regression model,*

$$\sigma(\boldsymbol{\beta}^T \tilde{\boldsymbol{x}}) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

   (a) [1pt] Show that this model satisfies

$$\frac{\sigma(\boldsymbol{\beta}^T \tilde{\boldsymbol{x}})}{1 - \sigma(\boldsymbol{\beta}^T \tilde{\boldsymbol{x}})} = e^{\beta_0 + \beta_1 x}$$

   (b) [3pts] Explain the difference between generative and discriminative classifiers. Is logistic regression a generative or a discriminative model ?

3. *[5pts] Consider a neural network with two hidden layers: $d = 4$ dimensional inputs, 2 units in the first hidden layer, 3 units in the second hidden layer and a single output.*

   a) Draw a picture of the network

   b) Write out an expression for $y(x)$ assuming ReLU activation functions. Be as explicit as possible.

   c) How many parameters are there?