Assignment 3

Homework assignments will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited. Electronic submission on Canvas is mandatory.

Submission Instructions You shall submit a zip file named Assignment3_LastName_FirstName.zip which contains:

- python files (.ipynb or .py) including all the code, comments, plots, and result analysis. You need to provide detailed comments in English.
- pdf file including (a) solutions for questions 1 and 2 (b) descriptions of observations for questions 3,4,5.

Table 1: Data set for question 2

data	x_{i1}	x_{i2}	y_i	α_i
$\mathbf{x}_1$	4	2.9	1	0.414
$\mathbf{x}_2$	4	4	1	0
$\mathbf{x}_3$	1	2.5	-1	0
$\mathbf{x}_4$	2.5	1	-1	0.018
$\mathbf{x}_5$	4.9	4.5	1	0
$\mathbf{x}_6$	1.9	1.9	-1	0
$\mathbf{x}_7$	3.5	4	1	0.018
$\mathbf{x}_8$	0.5	1.5	-1	0
$\mathbf{x}_9$	2	2.1	-1	0.414
$\mathbf{x}_{10}$	4.5	2.5	1	0

1. **Support Vector Machines** (20 points) Given 10 points in Table 1, along with their classes and their Lagranian multipliers (α_i), answer the following questions:
   (a) (7 pts) What is the equation of the SVM hyperplane $h(x)$? Draw the hyperplane with the 10 points.
   (b) (8 pts)What is the distance of x_6 from the hyperplane? Is it within the margin of the classifier?
   (c) (5 pts)Classify the point $z = (3, 3)^\top$ using $h(x)$ from above.

2. **Support Vector Machines** (20 points) The SVM loss function with slack variables can be viewed as:

$$\min_{\mathbf{w},b} \frac{||\mathbf{w}||^2}{2} + \gamma \sum_{i=1}^{N} \underbrace{\max(0, 1 - y_i f(\mathbf{x}_i))}_{\text{Hinge loss}} \tag{1}$$

The hinge loss provides a way of dealing with datasets that are not separable.

(a) (8 pts)Argue that $l = \max(0, 1 - y\mathbf{w}^\top \mathbf{x})$ is convex as a function of $\mathbf{w}$.

(b) (5 pts) Suppose that for some $\mathbf{w}$ we have a correct prediction of f with $\mathbf{x}_i$, i.e. $f(\mathbf{x}_i) = \text{sgn}(\mathbf{w}^\top \mathbf{x}_i)$. For binary classifications ($y_i = +1/-1$), what range of values can the hinge loss, l, take on this correctly classified example? Points which are classified correctly and which have non-zero hinge loss are referred to as margin mistakes.

(c) (7 pts) Let $M(\mathbf{w})$ be the number of mistakes made by $\mathbf{w}$ on our dataset (in terms of classification loss). Show that:

$$\frac{1}{n} M(\mathbf{w}) \leq \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i)$$

In other words, the average hinge loss on our dataset is an upper bound on the average number of mistakes we make on our dataset.

3. **Decision Trees** (20 points) Implement a Decision Tree model for the Titanic data set (only use the train file).

   - Explain how you preprocess the features.
   - Divide the train file into training data and testing data.
   - Build a tree on the training data and evaluate the classification performance on the test data.
   - Compare Gini index and Information Gain.
   - Report your best accuracy on the test data set.
   - Give a brief description of your observations.

4. **Boosting** (20 points) Implement AdaBoost for the Titanic data set. You can use package/tools to implement your decision tree classifiers. The fit function of DecisionTreeClassifier in sklearn has a parameter: sample weight, which you can use to weigh training examples differently during various rounds of AdaBoost.

   - Plot the train and test errors as a function of the number of rounds from 1 through 500.
   - Report your best accuracy on the test data set.
   - Give a brief description of your observations.

5. **Neural Networks** (20 points) Apply a Neural Network (NN) model to predict a handwritten digit images into 0 to 9. The pickled file represents a tuple of 3 lists : the training set, the validation set and the testing set. Each of the three lists is a pair formed from a list of images and a list of class labels for each of the images. An image is represented as numpy 1-dimensional array of 784 (28 x 28) float values between 0 and 1 (0 stands for black, 1 for white). The labels are numbers between 0 and 9 indicating which digit the image represents. The code block below shows how to load the dataset.

```
import cPickle, gzip, numpy

# Load the dataset
f = gzip.open('mnist.pkl.gz', 'rb')
train_set, valid_set, test_set = cPickle.load(f)
f.close()
```

- Plot the train, validation, and test errors as a function of the epoches.
- Report the best accuracy on the validation and test data sets.
- Apply early stopping using the validation set to avoid overfitting.
- Give a brief description of your observations.