Assignment 3

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) What is the equation of the SVM hyperplane $h(x)$? Draw the hyperplane with the 10 points.

   (b) What is the distance of x_6 from the hyperplane? Is it within the margin of the classifier?

   (c) Classify the point $z = (3,3)^T$ using $h(x)$ from above.

Table 1: Data set for question 1

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

2. **Support Vector Machines** (20 points) Create a binary (2 feature) dataset where the target (2-class) variable encodes the XOR function. Design and implement a SVM (with a suitable kernel) to learn a classifier for this dataset. For full credit, explain the kernel you selected, and the support vectors picked by the algorithm. Redo all the above with multiple settings involving more than 2 features. Ensure that your kernel is able to model XOR in all these dimensions. Now begin deleting the non-support vectors from your dataset and relearn the classifier. What do you observe? Does the margin increase or decrease? What will happen to the margin if the support vectors are removed from the dataset? Will the margin increase or decrease?(You can use packages/tools for implementing SVM classifiers.)

Table 2: Data for Question 3

Instance	a_1	a_2	a_3	Class
1	T	T	5.0	Y
2	T	T	7.0	Y
3	T	F	8.0	N
4	F	F	3.0	Y
5	F	T	7.0	N
6	F	T	4.0	N
7	F	F	5.0	N
8	T	F	6.0	Y
9	F	T	1.0	N

3. **Decision Trees** (20 points) Please use the data set in Table 2 to answer the following questions:

   (a) Show which attribute will be chosen at the root of the decision tree using information gain. Show all split points for all attributes. Please write down every step including the calculation of information gain of the attributes at each split.

   (b) What happens if we use **Instance** as another attribute? Do you think this attribute should be used for a decision in the tree?

4. **Boosting** (20 points) Implement AdaBoost for the Banana dataset with decision trees of depth 3 as the weak classifiers (also known as "base classifiers"). 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 10. Give a brief description of your observations.

5. **Neural Networks** (20 points) Develop a Neural Network (NN) model to predict class labels for the Iris data set. Report your training and testing accuracy from 5-fold cross validation. You can use packages such as TensorFlow.