DSC 255: Machine learning

Homework 6

Mathematical and conceptual exercises

1. Draw the decision boundary in $\mathbb{R}^2$ that corresponds to the prediction rule $\text{sign}(2x_1 - x_2 - 6)$. Make sure to clearly indicate where this boundary intersects the axes. Show which side of the boundary is classified as positive and which side as negative.

2. A particular labeled data of n points is randomly permuted and then the Perceptron algorithm is run on it, repeatedly cycling through the points until convergence. It converges after making k updates. For each of the following statements, say whether it is **definitely true** or **possibly false**, and give a brief reason.

   (a) The data set is linearly separable.

   (b) If the process were repeated with a different random permutation, it would again converge.

   (c) If the process were repeated with a different random permutation, it would again converge after making k updates.

   (d) k is at most n.

3. The Perceptron algorithm is run on a data set, and converges after performing $p+q$ updates. Of these updates, p are on data points whose label is -1 and q are on data points whose label is $+1$. What is the final value of the parameter b?

4. An SVM classifier is learned for a data set in $\mathbb{R}^2$. It is given by $w = (3, 4)$ and $b = -12$.

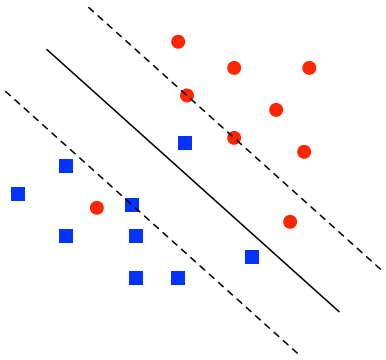
   (a) Draw the decision boundary, making sure to clearly indicate where it intersects the axes.

   (b) Draw the left- and right-hand boundaries, also clearly making where they intersect the axes.

   (c) What is the margin of this classifier?

   (d) How would the point $(2, 2)$ be classified?

5. The picture below shows the decision boundary obtained upon running soft-margin SVM on a small data set of blue squares and red circles.



(a) Copy this figure and mark the support vectors. For each, indicate the approximate value of the corresponding slack variable.

(b) Suppose the factor C in the soft-margin SVM optimization problem were increased. Would you expect the margin to increase or decrease?

6. The dual form of the Perceptron algorithm is used to learn a binary classifier, based on n training points. It converges after k updates, and returns a vector α and a number b. For each of the following statements, indicate whether it is **necessarily true** or **possibly false**, and give a brief justification.

   (a) Each α_i is either 0 or 1.

   (b) $\sum_i \alpha_i = k$.

   (c) α has at most k nonzero coordinates.

   (d) The training data must be linearly separable.

Programming exercises

7. *Binary Perceptron.* In this problem, you will code up the (binary) Perceptron algorithm and use it to classify the Iris data set.

   (a) Write code for two functions:

   - The first function takes as input parameters w, b of a linear classifier as well as a data point x, and returns the label for that point: $\text{sign}(w \cdot x + b)$. The label is either $+1$ or -1.
   - The second function takes as input an array of data points and an array of labels (where each label is $+1$ or -1), and runs the Perceptron algorithm to learn a linear classifier w, b. The algorithm should begin by randomly permuting the data points.

   In your writeup, give the code for these two functions.

   (b) Load in the Iris data set. You can do this by simply invoking:

```
from sklearn import datasets
iris = datasets.load_iris()
x = iris.data
y = iris.target
```

   The data has four features and three labels. Restrict it to features 1 and 3 (the second and fourth columns, sepal width and petal width) and to labels 0,1. Recode label 0 as -1, since this is what the Perceptron algorithm is expecting.

   (c) Now run the Perceptron algorithm on the data. In your writeup, show a plot with the data points (where the two labels have different colors) and the resulting decision boundary.

   (d) Now modify your code from part (a) to count the *number of updates* made by the Perceptron algorithm while it is learning. Run the algorithm 20 times and keep track of the number of updates needed each time. In your writeup, include a histogram of these values.

8. *Support vector machine.* As you did with the Perceptron, use the Iris data set, but this time use features 0 and 2, and labels 1,2.

   (a) Is this data linearly separable?

(b) Use `sklearn.svm.SVC` to fit a support vector machine classifier to the data. You will need to invoke the option `kernel='linear'`. Try at least 10 different values of the slack parameter C. In your writeup, include a table that shows these values of C and for each of them gives the training error and the number of support vectors.

(c) Which value of C do you think is best? For this value, include a plot of the data points and the linear decision boundary.