Notebook Assignment 2: - due date February 10

Linear Machines (updated November 14, 2024)

Consider the double semi-circle learning task for which an example dataset is shown in Fig. **??**. There are two semicircles of width `thk` and inner radius `rad`, separated by `sep` as shown with the -1 class being and the $+1$ class being blue. The dataset has been stored as a serialized object in a *pickle* file. You can use the following script to load this object for a dataset with `sep=5` and generates the scatter plot shown in Fig. **??**.

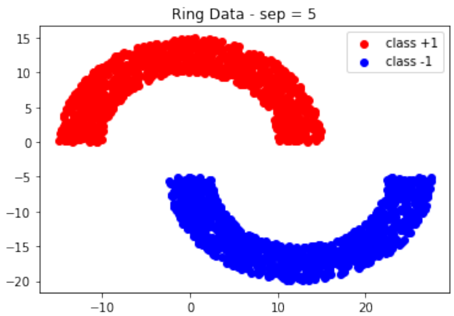

FIGURE 1. Problem 1

```
import numpy as np
import pandas as pd
import matploblib.pyplot as plt

filename = "data/ring_data_5.pkl"
df = pd.read_pickle(filename)
data = df.to_numpy()
ndat, m1 =  np.shape(data)
ndim       = n1-1
```

```
X = np.transpose(data[:,0:ndim])
Y = np.reshape(data[:,ndim],(ndat,1))

indx = np.where(Y==1)
plt.scatter(X[1,indx],X[2,indx],c='r')
indx = np.where(Y==-1)
plt.scatter(X[1,indx],X[2,indx],c='b')
plt.legend(('class +1','class -1'))
plt.title('Ring Data - sep = 5')
```

Let us consider a binary classification problem for a dataset $\mathcal{D} = \{(x_k, y_k)\}_{k=1}^N$ where $x_k \in \mathbf{R}^n$ is an *input sample* and $y_k \in \{-1, +1\}$ is a target (the kth sample's actual class). We are interested in determining a perceptron model, $h_w : \mathbb{R}^n \to \{-1, +1\}$ that takes values

$$h_w(x) = \text{sgn}(w^T x)$$

where $w \in \mathbb{R}^n$ is a weight vector and $x \in \mathbb{R}^n$ is a given input vector from the dataset. We are interested in finding a weight vector that minimizes the average classification error over the dataset. The loss function is

$$L(y, h_w(x)) = \mathbb{1}\left(y \neq h_w(x)\right)$$

so that $L(y, h_w(x)) = +1$ if the model's predicted classification matches that of the actual target, y and is -1 otherwise.

A recursive algorithm for training perceptron models is called the *perceptron learning algorithm* (PLA). The algorithm initializes the weight w to a random vector and then uses two data matrices $\mathbf{X} \in \mathbb{R}^{n \times N}$ and $\mathbf{Y} \in \mathbb{R}^N$ formed from the inputs and targets of the dataset $\mathcal{D}$

$$\mathbf{X} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_N \end{array} \right], \quad \mathbf{Y} = \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_N \end{array} \right]$$

The algorithm recursively updates the weight vector by selecting a sample (x_ℓ, y_ℓ) from the dataset. If that input x_ℓ is correctly classified by h_w,

$$h_w(x_\ell) = \text{sgn}(w^T x_\ell) = y_\ell$$

then the weight w is left unchanged. If x_ℓ is misclassified by h_w, then the algorithm updates the weight as

$$w \leftarrow w + \eta y_\ell x_\ell, \qquad \text{if } y_\ell w^T x_\ell < 0$$

This algorithm continues until all samples in the dataset are correctly classified by the model. If the dataset is linearly separable (i.e. there is a weight w such that h_w correctly classifies all data samples in the dataset), then the algorithm terminates after a finite number of updates.

Problem 1: For the dataset `ring_data_5.pkl` with `sep=5`, use the fixed increment perceptron training algorithm (PLA) to identify a perceptron that separates the data for a randomized initial weight. Determine the perceptron's classification error on the data set and the number of iterations the PLA used before stopping. Plot the data and the hyperplane characterizing the final model.

Problem 2: Let $\mathcal{D} = \{(x_k, y_k)\}_{k=1}^N$ denote the pickle dataset where x_k is the kth input and y_k is the kth target. Let

$$\widehat{R}(w \,|\, \mathcal{D}) = \frac{1}{N} \sum_{k=1}^{N} (y_k - h(x_k \,|\, w))^2$$

denote the empirical risk over dataset $\mathcal{D}$ for the perceptron model with weight vector w. Use the Moore-Penrose pseudoinverse to determine the weight vector minimizing the mean squared error in the empirical risk. Evaluate your classifier's classification error on the data set. Plot the data and the hyperplane characterizing your optimal model.

Problem 3: Repeat the first two parts using the dataset with a separation of -5 (`ring_data_-5.pkl`). This data set is *not* linearly separable. Determine the error rate and MSE for both algorithms. Discuss the difference that you see in using the PLA or Pseudoinverse for non-separable datasets.

Problem 4: For the data set, `ring_data_-5.pkl`, use the pseudoinverse method to find a perceptron based on the following 3rd order polynomial features

$$\phi(x_1, x_2) = \left[\begin{array}{cccccccccc} 1 & x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 & x_1^3 & x_1^2 x_2 & x_1 x_2^2 & x_2^3 \end{array}\right]$$

to minimize the MSE in the empirical risk over the dataset. Determine the classifier's classification error on the dataset. Plot the data and the hyper-surface characterizing your optimal model