Machine Learning (CS 6140)
Homework 4

Instructor: Ehsan Elhamifar
Due Date: December 12, 2017, 10:00am

1. Feed Forward Neural Network Implementation. Implement a Feed Forward Neural Network in Python, with an input layer with S_1 units, one hidden layer with S_2 units, and an output layer with S_3 units using the backpropagation algorithm. The network will be trained using data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$ for $\boldsymbol{x}_i \in \mathbb{R}^{S_1}$ and $\boldsymbol{y}_i \in \mathbb{R}^{S_3}$. The code must allow specifying the following activation functions: sigmoid, hyperbolic tangent and rectifier linear activation. The code must output all the learned weights and biasses of all layers as well as the activations of the last layer.

2. Auto-Encoder Implementation. Implement an auto-encoder neural network in Python, by modifying the code in question 1. The network receives the input data $\{\boldsymbol{x}_i\}_{i=1}^N$. The code must allow specifying the following activation functions: sigmoid, hyperbolic tangent and rectifier linear activation. The code must output all the learned weights and biasses of all layers as well as the activations of the last layer.

3. Testing Algorithms on Data.

Part A. Take the provided dataset, which consists of face images from 10 individuals corresponding to 10 classes.

a) Train the NN in the question 1 to separate the training face images in 10 classes. Report the classification error on the test data as a function of the number of neurons/units in the hidden layer. Report the results for all three activation functions. How does the result depend on activation function and the number of hidden neurons? Fully explain.

b) Train 1 versus all SVM classifier to separate the training face images in 10 classes. Report the classification error on the test data.

c) Train 1 versus all logistic regression classifier to separate the training face images in 10 classes. Report the classification error on the test data.

d) Repeat parts i, ii, iii by applying PCA to data first to reduce the dimension to $d = 100$ and then learning a classifier. How does the result change with respect to using the original data.

e) Repeat parts i, ii, iii by applying Auto Encode to data first to reduce the dimension to $d = 100$ and then learning a classifier. How does the result change with respect to using PCA and using the original data.

f) Visualized the data by applying PCA to data and reduce the dimension of data to $d = 2$. Repre-

sent different classes by a different marker and color in your plot. Are the data separated according to classes? Explain.

g) Apply Kmeans to the original dataset with $K = 10$. Show the results by plotting the 2-dimensional data and indicating data in each cluster by a different color. Report the clustering error on the dataset. Explain why Kmeans is or is not successful in recovering the true clustering/grouping of the data. Repeat this part by applying Kmeans to $d = 100$ dimensional PCA representations.

h) Apply Spectral Clustering with Gaussian RBF kernels to the original data. For spectral clustering, connect each point to its K nearest neighbors using Euclidean distance between points and for any two points i and j connected, set the weights to be $w_{ij} = \boldsymbol{e}^{-\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2/(2\sigma^2)}$. Try several values of K and σ and report the clustering error on the dataset as a function of K and σ. Explain why Spectral Clustering is or is not successful in recovering the true clustering/grouping of the data. How does spectral clustering perform compared to kmeans?

Homework Submission Instructions:
–Submission: You must submit all your plots and your Python code (.py file) via email, BY THE DEADLINE. To submit, please send an email to me and CC both TAs.
- The title of your email must be "CS6140: HW4: Your First and Last Name".
- You must attach a single zip file to your email that contains all python codes and plots and a readme file on how to run your files.
- The name of the zip file must be "HW4: Your First and Last Name".