ECE 805 – Machine Learning

Homework #04

Spring 2023

(due on Tuesday, 28 February 2023, 00:00, submitted via Teams)

General instructions

This assignment consists of two questions. You are requested to submit both your code and a report; the report should contain the requested visualisation plots and your answers to possible questions.
Please submit one compressed file (zip), named ECE805_HW4_Yourname.

Exercise 1 (50%):

This exercise is about Online Learning. Initially, you will be requested to train an Online Supervised Learning (OSL) model, and to gradually implement two Online Active Learning (OAL) strategies.

You are provided with the following files:

- **hw4_ex1_dataset.csv**: a simulated time-series dataset, corresponding to a classification problem. Each row corresponds to an arriving example; the last column being the true class.
- **hw4_ex1_classifier.csv**: it contains the Classifier class (a standard fully connected neural network).
- **hw4_ex1_main.csv**: it contains part of the code for online gradient descent.
- **hw4_ex1_auxiliary.csv**: it contains some auxiliary functions.
- **requirements.txt**: it provides the Python version and package versions. It is strongly suggested to use the same versions.

Familiarise yourself with these files. You will be requested to fill in the missing parts (displayed with "TODO") in the scripts hw4_ex1_classifier.csv and hw4_ex1_main.csv. Do not modify any other part of the code.

(a) Since the dataset corresponds to a multi-class classification problem, we will be using the "categorical cross-entropy" loss function. Complete all the missing code in hw4_ex1_classifier.csv, including, the appropriate output activation function.

    Also, in hw4_ex1_main.csv, complete the missing code up to (including) Line 60. For target encoding, use the Keras' built-in function "to_categorical".

(b) This question is about OSL. Complete Line 64 to incrementally train the classifier using the most recent example. Run the online gradient descent function using OSL twice, one with 1 epoch and the other with 10 epochs (Line 95). Provide the generated plot, describe and explain your findings.

(c) This question is about OAL and a **fixed** uncertainty sampling strategy. Complete Lines 70 and 73 to train the classifier (budget-permitting) when the most recent example's maximum prediction probability is less than or equal to a threshold. Do not forget to increment the budget spent. Run the algorithm with thresholds: 0.2, 0.4, 0.6, and 0.8. Provide the generated plot, describe and explain your findings.

(d) This question is about OAL and this **variable** uncertainty sampling strategy:

$$\theta_{new} = \theta_{old}(1 - s)\ if\ max.prediction\ prob \leq \theta_{old}$$
$$\theta_{new} = \theta_{old}(1 + s)\ if\ max.prediction\ prob > \theta_{old}$$

Use $s = 0.01$ and complete Line 76. Repeat the analogous experiment with initial thresholds: 0.2, 0.4, 0.6, and 0.8. Provide the generated plot, describe and explain your findings.

(e) Provide the generated plot which compares OSL with OAL variable uncertainty sampling with initial threshold 0.8 (Line 126). Discuss your findings.

Exercise 2 (50%):

It is not necessary to write code, but do submit any code that you might use along with the handwritten assignment. Write down all your reasoning.

(a) Cluster the given 8 data points into K = 3 clusters using the K-means algorithm and Euclidean distance. The coordinates of the data points [x1, x2] are:

id	X1	X2
1	2	8
2	2	5
3	1	2
4	5	8
5	7	3
6	6	4
7	8	4
8	4	7

i) Calculate the distance matrix of the points.

ii) Assume that you initialize K-means with K randomly chosen data points. Let's suppose that points 3, 4 and 6 are selected. After one iteration of the K-means algorithm, please provide the coordinates of the resulting centroids. Given that you have a new data point [4,5] to which cluster will it be assigned.

(b) Suppose that we apply PCA to a dataset with two dimensions, and the resulting eigenvalues are identical, what does this imply about the significance of the dimensions? Given we may have a million samples would it be a good choice to pursue dimensionality reduction? Provide an explanation and depict a dataset where the eigenvalues are of the same magnitude.

(c) Consider we apply PCA on a two-dimensional data set and we get the eigenvalue 6 and 2. What would happen if you pursuit dimensionality reduction on this dataset? Now for a three-dimensional dataset we get eigenvalues 0, 1, and 0. Explain what these values mean.