

Homework No. 4
Due 03/23 (11:59pm), 2022
150 points

Objectives

1. Apply Kernel SVM and MLP classification algorithms to the *fashion-MNIST* dataset
2. Use *k*-fold cross validation to identify the best way to rescale and preprocess the data
3. Use *k*-fold cross validation to identify the parameters that optimize performance (generalization) for each method
4. Compare the accuracy and identify correlation between the outputs of the two methods

Problem

For this homework, you will apply the following classification methods to the *fashion-MNIST* classification data

1. Kernel Support Vector Machines
 2. Multilayer Perceptrons
- Apply 4-fold cross-validation to the provided training data subset to train your classifiers and identify their *optimal parameters*. In addition to the classifier's parameters (e.g. regularization, kernel, Number of layers/nodes, learning rate, etc.), you should also consider the following 4 ways to preprocess and rescale the data:
 - a) No preprocessing
 - b) StandardScaler
 - c) RobustScaler
 - d) MinMaxScaler
 - After fixing the classifiers' parameters, apply each method to the provided testing data subset to predict and analyze your results. *Compare the accuracy* obtained during training (average of the cross-validation folds) to those of the test data and comment on the results (overfitting, underfitting, etc.)
 - Analyze the correlation between the output of the 2 classifiers by displaying the *predict_proba* of SVM vs. *predict_proba* of MLP (using test data). Using these scatter plots (one per class), identify (if available) the following 3 groups
 - G-1: Samples that are easy to classify correctly by the SVM, but hard to classify by MLP
 - G-2: Samples that are easy to classify correctly by the MLP, but hard to classify by SVM
 - G-3: Samples that are hard to classify correctly by both methodsFor each group, display few samples (as images) and identify any common features among them.

What to submit?

- A report that
 - **Describes** your experiments, the parameters considered for each method, etc.
 - **Summarizes, explains** (using concepts covered in lectures) and **compares** the results (using plots, tables, figures)
- Do not submit your source code
- Your report needs to be a single file (MS Word or PDF)
- Your report cannot exceed 10 pages using a font of 12
- Assign numbers to all your figures/tables/plots and use these numbers to reference them in your discussion