CSCM45 Big Data and Machine Learning

Coursework 2: Object Recognition

2 Pages max

**Introduction**

Provide an overview of the problem, your proposed solution, and your experimental results.

With the provided dataset being so big, with many features, it is potentially going to take a long time to process and train the Neural Networks. With using a dimensionality reduction techniques along side a NN, it will be compared to see if the reduction in potential accuracy is a valid trade off for processing time.

The aim of the project is to see what method creates the best prediction accuracy form the provided dataset. This will be using LDA, PCA, SVM and a Neural Network to see what provides the best results. The accuracy, as a percentage, and how long the algorithms take to run will be used as factors/benchmarks for comparison.

**Method**

Present your proposed method in detail. This should cover how the features are extracted, any feature processing you use (e.g. clustering and histogram generation, dimensionality reduction), which classifier(s) is/are used, and how they are trained and tested. This section may contain multiple sub-sections.

**Extracting the data from the Dataset**

The first task will be using the provided function to extract the features from the dataset and then add them to a numpy array. The same will be done to the training data set provided. Training and testing data labels will be used for classifying the data and testing the accuracy.

**Dimensionality Reduction Techniques**

Both PCA and LDA will be used. First by themselves, fitting and then testing the data. Then by fitting the data with the output results being fed through a Neural Network.

**Non Dimensionality Techniques**

The data will also be fed through a NN without any reduction to see a comparison, along with the running time of the task processing. As well as SVM, to have a comparison of non-dimensionality techniques.

**Results**

Present your experimental results in this section. Explain the evaluation metric(s) you use and present the quantitative results (including the confusion matrix).

**Conclusion**

Provide a summary for your method and the results. Also, provide your critical analysis; that is the shortcomings of your method and how they may be improved.

**References**

Include references where appropriate. References are not included in the page limit.

Submission

Submit your work electronically to Blackboard. Your report should be in PDF format only. Compress your Python source code and report into a Single Zip file. The deadline for this coursework is 11AM Monday 9th December.