



Machine Learning Online | Assignment - IV

Fashion MNIST | Classification Problem

1. Download the Fashion MNIST Dataset from [Kaggle \(https://www.kaggle.com/zalando-research/fashionmnist/\)](https://www.kaggle.com/zalando-research/fashionmnist/). Visualise the dataset using matplotlib. Report the number of examples and features and no of classes.
2. Take the first 40,000 examples from the data. Split the dataset into train and test set (80:20 Ratio) and train classifiers as given below.
3. We have learned that Logistic Regression is a Binary Classifier, but the current problem is an example of [Multiclass Classification \(https://en.wikipedia.org/wiki/Multiclass_classification\)](https://en.wikipedia.org/wiki/Multiclass_classification) - and we use generally use One Vs Rest or One vs One schemes to do such a classification. Define a Logistic Classifier class and train your model using **One vs One Scheme** to give predictions on the test data.
 - **Note** : In the one-vs.-one reduction, one trains $K(K - 1) / 2$ binary classifiers for a K-way multiclass problem; each receives the samples of a pair of classes from the original training set, and must learn to distinguish these two classes. At prediction time, a voting scheme is applied: all $K(K - 1) / 2$ classifiers are applied to an unseen sample and the class that got the highest number of "+1" predictions gets predicted by the combined classifier.
4. Repeat part-3 using Sci-kit Learn [Logistic Regression Class \(http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html\)](http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). Which scheme does the algorithm use - One Vs Rest or One Vs One scheme use? Compare the results when you use - one vs one scheme and one vs rest scheme. Have a look at multi_class parameter in the library documentation. Write your predictions on the test data in '.CSV' file.
5. Repeat the classification task using K-NN Algorithm to classify the test data. Since it might take time, so you can reduce the number of training and test data samples.

(Upload your Jupyter-Notebooks to Github, your personal repository)