Task 2

DEADLINE: 8:30 am Friday, 7th March, 2020

No late submissions would be entertained

Do NOT plagiarize!

The goal in this assignment is to develop a machine learning model that can classify between images of T-shirts and dress-shirts.

You are given the following files:

- **TrainData.csv**: It contains 12000 training examples. Each row contains 784 values. The dataset has been derived from Fashion-MNIST dataset. Each example is a flattened 28x28 pixel gray-scale image. You can reshape the examples to visualize what each image looks like.
- **TrainLabels.csv**: This file contains true labels for the examples in TrainExamples.csv
- **TestData.csv**: This file contains test examples.
- You can load train and test data using the following code:

```
Xtr=np.loadtxt("TrainData.csv")
Ytr=np.loadtxt("TrainLabels.csv")
Xts=np.loadtxt("TestData.csv")
```

- To visualize an example (say trainining example at index 10, you can use the following code):

```
import matplotlib.pyplot as plt
plt.imshow(Xtr[10].reshape([28,28]))
```

Tasks:

- Write a method named *ExtractFeatures*. It should take as arguments the examples, compute and return a feature vector, that you think would help in performing the task, corresponding to each example.
- You are supposed to train a model using your extracted features. Try at least 2 different classification techniques of your choice.
- Using 5-fold cross-validation, optimize hyperparameters for the models. Since the dataset is balanced, you can use classification accuracy as the performance metric.
- After choosing the best hyperparameters, use the complete training dataset to train the final model. Dump the model in the file named "myModel.pkl". You can take help from https://scikit-learn.org/stable/modules/model_persistence.html for saving your model.
- Create a file test.py. It should read the test examples from TestExamples.csv. **DO NOT CHANGE THE ORDER OF EXAMPLES**. Generate predictions for the test examples and save them to a csv file named "myPredictions.csv". You can use the following code to save your results:

```
np.savetxt("myPredictions.csv", Yts)
```

You have to submit the following files:

- Files containing the cross-validation code used for hyperparameter selection.
- Training.py: The file containing the code to train and dump the final model using the chosen hyperparameters.
- Testing.py: The file in which you load the saved model and generate and save predictions for test data.
- myPredictions.csv: The csv file containing your predicted labels for the **test set.**

- A pdf report that contains the following information:
    - What features did you use and why?
    - What classification techniques did you try?
    - Which of the methods (and for what hyperparameters) showed best cross-validation performance.
    - What test accuracy are you expecting?

Note:

Please name your submission files in the following manner:

<roll no>_<name>_<section>

For example: i18456_Ahmad_Ali_MLDS_A