Practical Machine Learning
ENGR 491/891

Programming Assignment 3

Spring 2022

Logistic Regression

ENGR 891: 100 points

Last Name 1: Liew

First Name 1: Chia Sin

NUID 1: 53463291

Obtained Score:

a) **Part A:** Optimal set of features for the Logistic Regression model.

Models that used all features performed the best for all trained models.

b) **Part A:** Using the table (given in Part A) report training accuracy, test accuracy, weighted average of the test precision, recall, F1 scores, degree of the optimal polynomial model (experiments 3), and optimal values of the model hyperparameters.

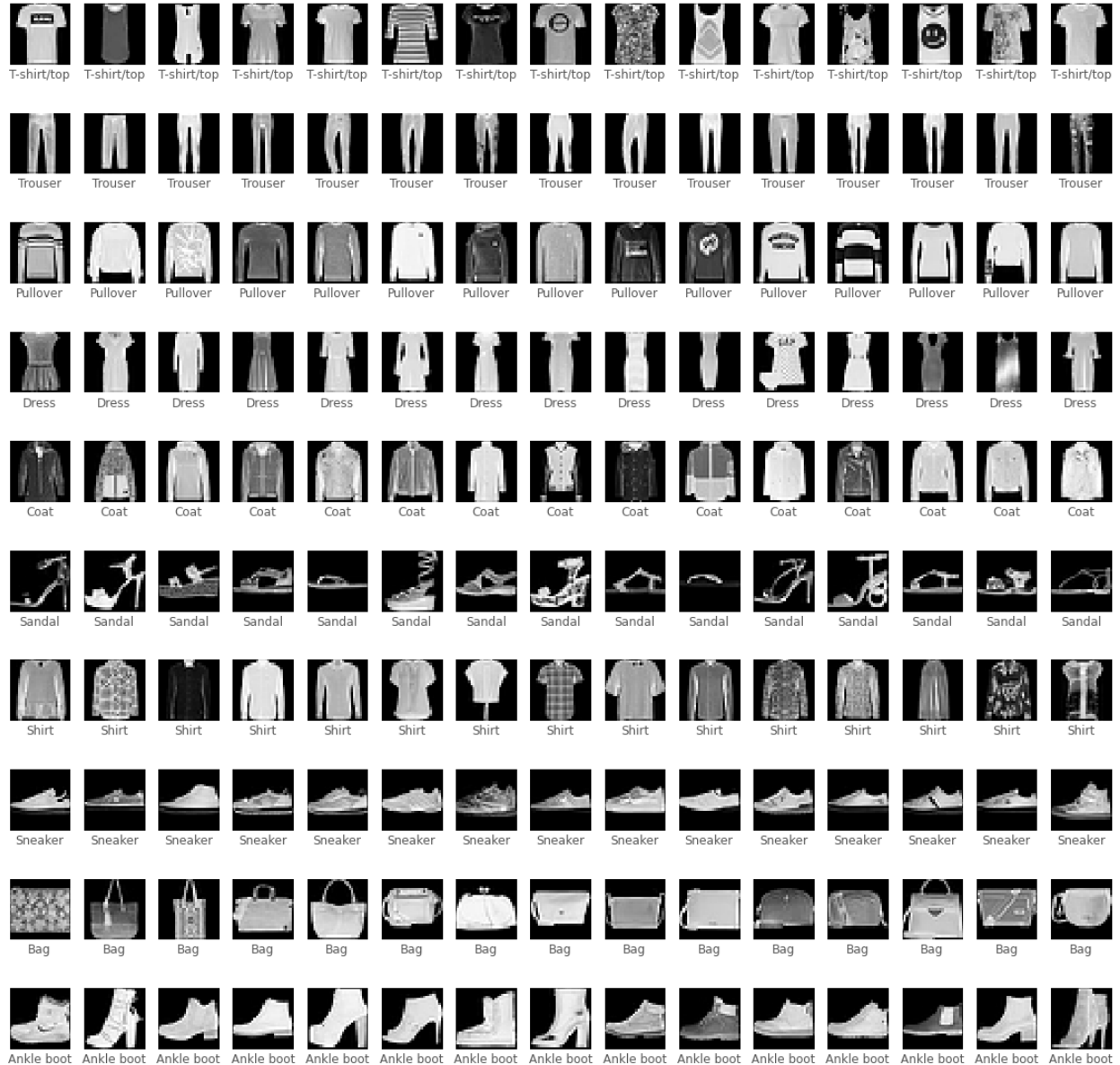|	train accuracy	Test accuracy	Weight precision	Weight recall	f1-score	Optimal hyperparameters	Wall time
KNN	1	0.68	0.7	0.68	0.67	{'n_neighbors': 25, 'p': 1, 'weights': 'distance'}	4min 56s
Logistic Regression (OvA)	0.54	0.52	0.49	0.52	0.46	{'C': 0.1, 'max_iter': 500, 'multi_class': 'ovr', 'solver': 'liblinear', 'tol': 0.001}	6.33s
Logistic Regression (Softmax)	0.54	0.52	0.52	0.52	0.48	{'C': 0.1, 'max_iter': 500, 'multi_class': 'multinomial', 'solver': 'newton-cg', 'tol': 0.001}	22.1s
Polynomial Logistic Regression	0.84	0.6	0.6	0.6	0.6	{'log_reg__C': 0.1, 'log_reg__max_iter': 1000, 'log_reg__multi_class': 'multinomial', 'log_reg__solver': 'lbfgs', 'log_reg__tol': 0.001, 'poly__degree': 4}	5h 54min 22s

c) **Part A:** Answers to Q-1 to Q-2.

- **Q-1)** Which model did you find most **effective** (optimal test performance)? Explain why this model performed better than other models in your experiments.

  KNN, using all features, is the most effective according to its test performace (Train accuracy of 1, test accuracy of 0.68, weighted precision of 0.7, weighted recall of 0.68 and f1-score of 0.67. I am not sure why KNN performs the best, but my best guess is that it is the only non-parametric model among the models we tried and since the dataset is relatively small and an instance based model works better in this case.

- **Q-2)** Which model did you find most **efficient** (less time for hyperparameter tuning)? Explain why it is efficient.

The most efficient model is Logistic Regression (Ova). It is the fastest because it is a linear regression model and there are comparatively less parameters to train compared to Logistic Regression (Softmax) and Polynomial Logistic Regression. It is also faster since it is model based instead of instance based like KNN.

d) **Part B:** Display 15 random images from each class.

e) **Part B:** Report train accuracy, test accuracy, and test classification report for experiment 4.

Train accuracy: 0.87
Test Accuracy: 0.85

```
Classification Report:
             precision    recall  f1-score   support

          0       0.81      0.81      0.81      1000
          1       0.98      0.95      0.97      1000
          2       0.73      0.74      0.73      1000
          3       0.83      0.87      0.85      1000
          4       0.74      0.76      0.75      1000
          5       0.94      0.92      0.93      1000
          6       0.63      0.57      0.60      1000
          7       0.91      0.94      0.92      1000
          8       0.93      0.94      0.94      1000
          9       0.95      0.94      0.95      1000

   accuracy                           0.85     10000
  macro avg       0.84      0.85      0.84     10000
weighted avg      0.84      0.85      0.84     10000
```
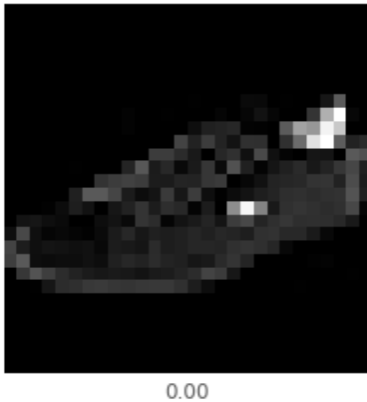
f) **Part B:** Answers to Q-3.

Sandal

0.00

This is a misclassified test image of a sneaker. The model misclassified this image as a sandal because of its color of black and the background also being black. Because of the high homogeny in color, the shoe laces were mistaken as some of the sandal straps.

0.00

This is a misclassified test image of a T-shirt/top. The model misclassified this image as a sandal because of reasons similar to the first image. The colors of the top and background are similar, and the model has a hard time to classify it and misclassify it as sandal because it has some strips on it.

Sneaker

0.00

This is a misclassified test image of an ankle boot. The model misclassified this image as a sneaker because as we can see, it looks like a high ankle sneaker and it is hard to tell from the image that it is actually a boot.

0.00
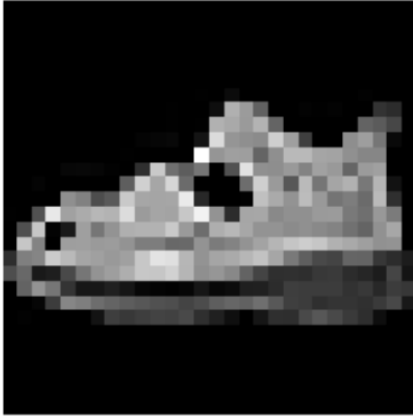
This is a misclassified test image of a sandal. The model misclassified this image as a sneaker because of the low resolution of the image, it is hard to tell if it is a sandal or some kind of shoes with laces.