Machine Learning – Kernel SVM Assignment

Name: Nour Mohamed Hussein Mahmoud Mohamed Kamaly

ID: 20191700701

Department: Scientific Computing Level 3

Section: 5

Feature selection:

- According to LaGrange interpolation techniques for finding a unique polynomial that passes through the observations, if the number of features (estimators/predictors) was more than the number of observations (data points), then the predictive line is not unique and the model is prone to overfitting, and this is the case here, 3780 features and 2000 observations, so some features had to be removed
- Two methods were used
    - Selecting k best features: K = 10
    - Hypothesis testing using p value:
      My null hypothesis is that the model will make use of the feature otherwise the label is independent on that feature.
      So, a p value < 0.05 proves the null hypothesis is true and a p value larger than it proves otherwise. 1754 features were selected in the end. This improved the model's performance.

Experiments:

- 6 different models were chosen
    - Linear SVM
    - SVM with linear kernel
    - SVM with RBF kernel
    - SVM with polynomial kernel
    - Stacking Classifier that takes the above 4 models and stacks them according to best performing to worst performing, 10 folds (cross validation), and logistic regression (2 classes only) as a final estimator
    - Convolution neural network (due to the small number of training set the model resulted in overfitting)
      2 convolution layers and 2 dense layers

Optimizer: stochastic gradient descent (learning rate = 0.01 & momentum = 0.8 – used high momentum value as the model used to be stuck at a local minimum {0.5})
Loss function: binary cross entropy
Batch size = 32

Pvalue models' results:

Model	Training accuracy	Testing accuracy
Linear SVM	0.976	0.725
Linear kernel function	0.9185	0.73
RBF kernel function	0.9485	0.725
Polynomial kernel function	1.0	0.735
Stacking Classifier	0.986	0.725

SelectKBest models' results:

Model	Training accuracy	Testing accuracy
Linear SVM	1.0	0.675
Linear kernel function	0.997	0.655
RBF kernel function	0.965	0.745
Polynomial kernel function	1.0	0.705
Stacking Classifier	1.0	0.675

CNN model results:

Model	Training accuracy	Testing accuracy
CNN	0.9803	0.5