***Student: Victor Pugliese***

***R#: 11492336***

***Course: Independent Study on Python Machine Learning for Petroleum Engineering Application (PETR 5000)***

***Self-Homework #5***

1. ***What is the fundamental idea behind Support Vector Machines?***

The fundamental idea behind Support Vector Machines is to define a decision boundary between two classes with the widest margin possible. It is possible that a few instances stay inside the margin.

1. ***What is a support vector?***

A support vector is any instance in the “street”, i.e. between the margin lines, that affects or determines the decision boundary. The decision boundary is fully supported by the instances located on the street.

1. ***Why is it important to scale the inputs when using SVMs?***

SVMs are sensitive to the feature scales. The algorithm will tend to ignore features with small values.

1. ***Can an SVM classifier output a confidence score when it classifies an instance? What about a probability?***

SVM classifier can output a distance measure that could be used a score, however SVM classifiers do not output probabilities for each class.

1. ***Should you use the primal or the dual form of the SVM problem to train a model on a training set with millions of instances and hundreds of features?***

The dual form of the SVM problem is proportional to a number between and . With millions of instances, the solver for the dual form will be too slow. The kernelized algorithms use the dual form, then in this case is better to use the LinearSVC or LinearSVR.

1. ***Say you trained an SVM classifier with an RBF kernel. It seems to underfit the training set: should you increase or decrease γ (gamma)? What about C?***

If the model underfits the training set, we should increase and/or C.

1. ***How should you set the QP parameters (H, f, A, and b) to solve the soft margin linear SVM classifier problem using an off-the-shelf QP solver?***
2. ***Train a LinearSVC on a linearly separable dataset. Then train an SVC and a SGDClassifier on the same dataset. See if you can get them to produce roughly the same model.***

See file: HML\_Chap05\_Exercise\_08.py

1. ***Train an SVM classifier on the MNIST dataset. Since SVM classifiers are binary classifiers, you will need to use one-versus-all to classify all 10 digits. You may want to tune the hyperparameters using small validation sets to speed up the process. What accuracy can you reach?***

See file: HML\_Chap05\_Exercise\_09.py

1. ***Train an SVM regressor on the California housing dataset.***