PA-2 REPORT-SVM

**Name**:Abhay Singh

**ID**: 1001669333

**- Describe SVM method**

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_1.png)

**- What is a support vector?**

Support vectors are the data points that lie closest to the decision surface (or hyperplane) • They are the data points most difficult to classify • They have direct bearing on the optimum location of the decision surface. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

**- What is the difference between One-vs-One and One-vs-All method**

The difference is the number of classifiers you have to learn, which strongly correlates with the decision boundary they create.

Assume you have NN different classes. One vs all will train one classifier per class in total Nclassifiers. For class ii it will assume ii-labels as positive and the rest as negative. This often leads to imbalanced datasets meaning generic SVM might not work, but still there are some workarounds.

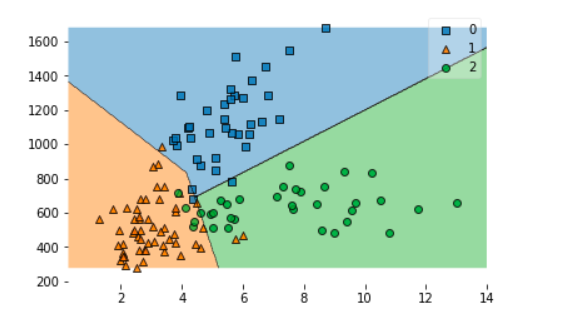
In one vs one you have to train a separate classifier for each different pair of labels. This leads to N(N−1)2N(N−1)2 classifiers. This is much less sensitive to the problems of imbalanced datasets but is much more computationally expensive.

**- explain what your criteria was for selecting the two attributes.**

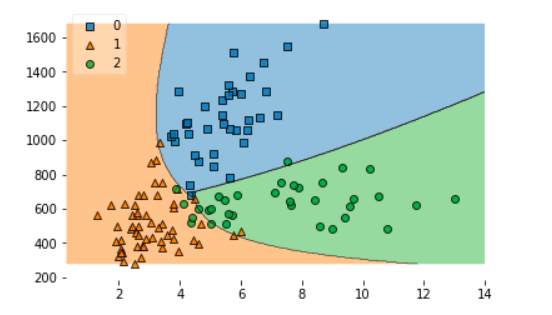
Univariate feature selection works by selecting the best features based on univariate statistical tests. Select features according to the k highest scores. It removes all but the k highest scoring features. For classification: [**chi2**](http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2), **[f\_classif](http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html" \l "sklearn.feature_selection.f_classif" \o "sklearn.feature_selection.f_classif)**, **[mutual\_info\_classif](http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html" \l "sklearn.feature_selection.mutual_info_classif" \o "sklearn.feature_selection.mutual_info_classif)**. Chi-squared stats of non-negative features for classification tasks. This score can be used to select the n\_features features with the highest values for the test chi-squared statistic from X, which must contain only non-negative features such as Booleans or frequencies. A chi square statistic is a measurement of how expectations compare to results.

**- Visualize SVM in a 2D projection for linear, and non-linear kernels.**

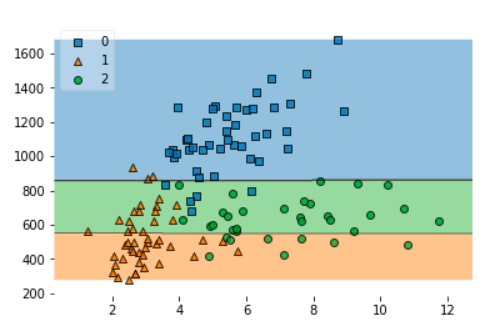
SVM plot for Linear kernel.

****

SVM plot for Non-Linear kernel, polynomial.

****

SVM plot for Non-Linear kernel, gaussian RBF

.

**- Interprete and compare the results.**

* From plots, non-Linear kernel gives more precise information if points are convoluted.
* Use linear kernel when number of features is larger than number of observations.
* Use non-linear kernel when number of observations is larger than number of features.
* Always try linear first since it is way faster to train and test.
* Linear kernel is best if the space between the observation is high. For the non-linear kernel, we will have to increase the dimensions for seraration.

**Reference**:

<https://stats.stackexchange.com/questions/91091/one-vs-all-and-one-vs-one-in-svm>

<http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2>

<https://www.investopedia.com/terms/c/chi-square-statistic.asp>