Lab 7 – SVMs

MACHINE LEARNING

SAUMAY AGRAWAL

16BCE1151

EXPERIMENT

- Please try SVM for your dataset.
- Try a pipeline of SVM parameters, C, Gamma, Kernel and report your observation
- Try SVM for Regression and tune the parameters. Choose an appropriate dataset for this.
- Explore the interactive SVM demo on https://dash-gallery.plotly.host/dash-svm
- and report your observations

ALGORITHM

SUPPORT VECTOR MACHINES

"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 (look at the below snapshot). Meanwhile, when used in regression, the SVM tries to find a hyperplane which tries to minimize the error in predictions (which is a difference of real values here, as opposed to categorical values in classification).

KERNEL TRICK IN SVM

SVM has a technique called the kernel trick. These are functions which takes low dimensional input space and transform it to a higher dimensional space i.e. it converts non separable problem to separable problem. These functions are called kernels. It is mostly useful in non-linear separation problem. There are various kernels like linear kernel, radial basis kernel, polynomial kernel and sigmoid kernel.

OBSERVATIONS

IRIS DATASET

- SVC observations
  - Attributes ‘sepal length’ and ‘sepal width’ were chosen for this experiement. Also, rows with target labels as ‘iris versicolor’ were also removed for binary classification.
  - For ‘linear’ kernel, maximum accuracy of 100% was achieved for penalty parameter C=1.0.
  - The average score of ‘rbf’ was found to be higher than that of ‘poly’.
  - For ‘sigmoid’ kernel, we get the second highest average score.

- SVR observations
  - For regression, attributes ‘petal length’ and ‘petal width’ were chosen from the same dataset.
  - The score is the coefficient of determination of predications in SVR. The best possible score is 1.0. It can be negative also for very arbitrary model).
  - Only ‘linear’ and ‘rbf’ kernel models performed well, while ‘poly’ and ‘sigmoid’ models gave very low accuracy.

OLYMPICS DATASET

- SVC observations
  - The dataset was subsampled for the year 2016 and fewer attributes than the original dataset. This was done to observe the variations by altering gamma and penalty parameter in less computations.
  - The attributes ‘Sport’ and ‘Team’(which represents country here) were chosen to classify whether an athlete won ‘Silver’ medal or not.
  - The ‘poly’ kernel outperformed ‘linear’ and ‘rbf’ kernels. ‘Sigmoid’ kernel again gave the lowest performance of all.

- SVR observations
  - ‘Height’ and ‘weight’ from the same dataset were used for regression.
  - ‘rbf’ kernel gave the best results, while ‘linear’ performed fairly well, as the data isn’t linearly separable here.
  - ‘poly’ and ‘sigmoid’ kernel models gave unacceptable results.

INFERENCES

Following were found from the results on datasets above and SVM Explorer by Dash:

- Radial Basis Kernel or ‘rbf’ is more versatile, as it performed well on all kinds of datasets (even on moons sample)
- Linear kernel is generally suitable for and limited to linearly separable data only.
- Polynomial kernel is generally computationally intensive with low performance , and sigmoid kernel’s performance is totally subjected to the dataset used.
- Gamma values monitor how well the decision boundary or hyperplane will fit the data. Lower gamma values will lead to too much generalisation or underfitting, while higher gamma values will lead to overfitting. Through experimentation, appropriate value of gamma should be identified for a particular dataset.
- Lower values of penalty parameter C will train the model less strictly, resulting in wider margins and less accuracy. Meanwhile, higher values will result in strict training of model, so that model is able to classify with more precision and gives smaller margins around the decision boundaries.