Challenge 5: Support Vector Machine

Neehar Yalamarti

Support Vector Machines are generally used for classification tasks where a hyper plane is used to separate two or more classes in the feature space. A commonly used kernel, also applied for the below task, is the Radial Basis Function kernel which follows the equation:

$$K(\mathbf{x}, \mathbf{x'}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{x'}||^2}{2\sigma^2}\right)$$

Here, $||x - x'||^2$ is the squared Euclidian distance between the two feature vectors. σ is a parameter where

$$\gamma = \frac{1}{2\sigma^2}$$

leading to:

$$K(x, x') = \exp(-\gamma ||x - x'||^2)$$

γ intuitively defines how far the influence of a single training sample reaches, a smaller value indicates encapsulation of a bigger region while a larger value indicates a smaller surrounding region. In the below case, an appropriate value of $\gamma = 1$ is chosen as a good influence of the training samples.

Another parameter used is the cost C which is a regularization parameter i.e. it is the penalty associated to the instances which are either misclassified or violates the maximal margin. Higher the value of C, fewer the misclassifications while a lower C results in a larger margin, reducing training accuracy. For the below case, an appropriate value of C = 0.15 is chosen to ovoid over fitting the training data.

The accuracy using the following parameters is 88% signifying a decent fit for the training data.
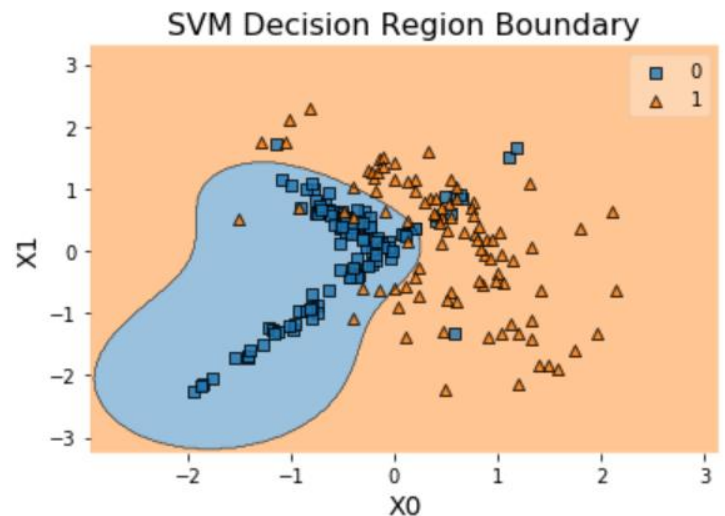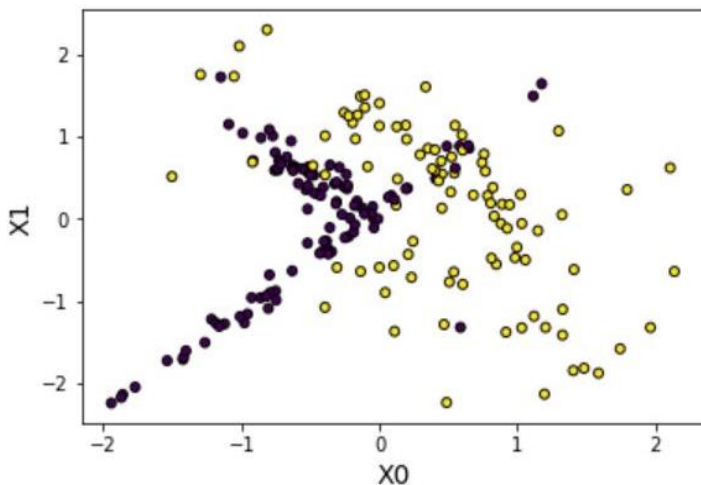
REFERENCES

[1] Scikit-learn.org. (2018). RBF SVM parameters — scikit-learn 0.20.0 documentation.
[2] Support Vector Machine – Wikipedia.org