Understanding Support Vector Machine Kernels Through Visualisation

Table of Contents

1. Introduction

GitHub Link- https://github.com/Sweety-gif-sys/Machine-Learning-Tutorial-

Support Vector Machines (SVMs) represent an effective type of supervised machine learning algorithm commonly applied to classification and regression. Their main attributes that endear them are a firm theoretical basis, resistance to overfitting in high dimensions, and performance on intermediate-sized datasets. Nonetheless, the most significant and possibly misinterpreted issue about SVMs is the role that kernel functions play.

Kernels help SVMs to capture non-linear complex relationships by implicitly projecting data to the feature space of higher dimensions. The kernel used can have a tremendous affect on the model behaviour, decision boundaries and predictive performance. Consequently, to be able to use SVMs in the real world, it is necessary to know how they operate as a practitioner (Shafi et al., 2022).

The tutorial concerns the influence of various SVM kernels on the classification and decision boundaries. We test a synthetic non-linearly separable dataset between linear, polynomial and radial basis function (RBF) kernels. It is visualised throughout to create intuition, thus the tutor is friendly to the learners new to kernel-based methods, but suitable enough to the advanced learners.

2. Overview of Support Vector Machines

As a concept, an SVM aims at identifying a decision boundary that will distinguish between data points representing different classes with the greatest margin. The distance between each separating hyperplane and the nearest data examples in each of the classes, or support vectors, is called the margin (Kok et al., 2021).

When the data is linearly separable, then such a hyperplane is obtainable in the feature space of input itself. But numerous real world data cannot be separated in a line, a straightforward linear decision boundary is inadequate. This restriction drives the invention of kernel techniques.
Mathematically, an SVM is an optimisation problem which trades off two conflicting goals:

1. Maximising the margin between classes.
2. Minimising classification errors.

The regularisation parameter C controls this balance determining the strictness with which the model penalises misclassified points. A low C gives a larger allowance but can withstand more misclassifications whereas when C is large, correct classification is of greater priority but it would overfit.

3. The Kernel Trick

The kernel trick is one of the main concepts that allow efficient handling of non-linear decision boundary by the SVMs. Rather than taking a transformation step to data into a higher- dimensional feature space, kernels directly compute inner-products of data points in this space (Anand et al., 2021).

Formally, a kernel function $K(x, x')$computes the dot product of transformed features:

$$K(x, x') = \phi(x) \cdot \phi(x')$$

where $\phi(\cdot)$represents a (possibly high-dimensional) feature mapping.

This has the advantage of avoiding the computational expense of directly implementing the transformation, and SVMs can work in very high or even infinite-dimensional spaces.

Common kernel functions include:

- Linear kernel
- Polynomial kernel
- Radial Basis Function (RBF) kernel

Each kernel induces a different geometry in the feature space, leading to distinct decision boundaries.

4. Dataset Description

In this study, a synthetic two-moons data set was to be used to depict the influence of various Support Vector machine kernels. The data has two overlapping half-circles and it is purposefully not linearly separable, hence suitable in assessing the shortcomings of linear classifiers and the benefits of kernel-based classifiers (Su et al., 2021).

The dataset has been produced through the make moons function of scikit-learn library. Gaussian noise was applied to give it some variation and more realistic to the flaws usually present in real world data. This makes the classification process non-trivial, and it offers a realistic assessment environment of various kernels.

The data was split into a training and testing data set with a 70/30 split, and thus, the ability of the model to perform in unseen data was evaluated and the chances of overfitting diminished. The two-dimensional representation of the data also makes the demarcation of the decision boundaries easy to visualise to facilitate intuitive interpretation and efficient comparison of the different types of kernels.

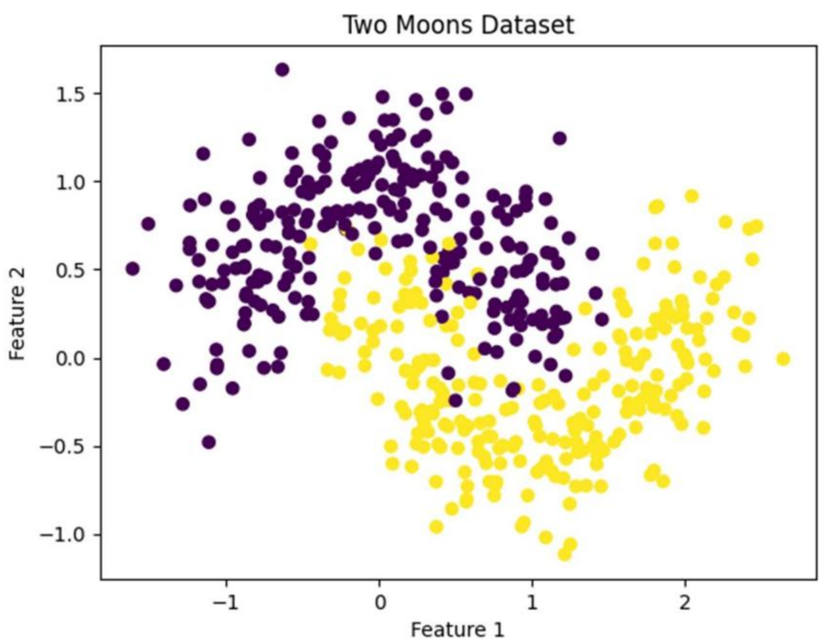

Figure 1: Scatter plot of the two-moons dataset showing non-linear class separation

5. Linear Kernel SVM

The simplest type of Support Vector Machine is the linear kernel, in which the classification is done by fitting a straight-line decision margin in the original feature space. This can be used to solve the problem when the data is separable linearly as an objective is to maximise the gap between classes and have a simple and interpretable structure of the model.

TWorking with the two-moons data, the linear SVM has difficulties in capturing the model data distribution. The curved nature and interweaving nature of the two classes necessitates a single linear boundary to provide effective separation. It subsequently leads to a large percentage of misclassification of data especially where there is overlap between data classes (Mastouri et al., 2021).

The model had test accuracy of about 85 which means that there is underfitting and over-biasing. Although linear kernels are computationally efficient and they can be interpreted with a lot of strength, this experiment indicates that they fail with non-linearly separable datasets and that more versatile kernel functions are needed in this case.

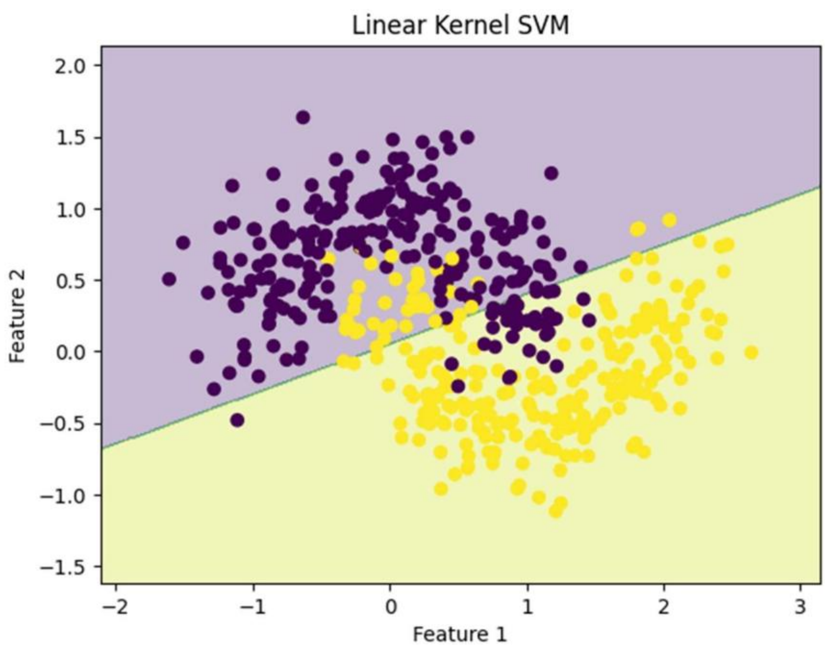

Figure 2: : Decision boundary produced by a linear kernel SVM

6. Polynomial Kernel SVM

The polynomial kernel optimally complements the functions of the linear Support Vector Machine by allowing non-linear interactions of features to a defined polynomial degree. This enables the model to capture curved and more complicated decision boundaries whilst still having a structure in terms of mathematics. The degree parameter can be used to ensure that the level of non-linearity is introduced into the classification process by increasing or decreasing the degree of the polygraph kernel.

On the two-moons data a three-degree polynomial kernel was used in this tutorial. Relative to the linear kernel, this method generates a more shapeless decision boundary that is more compliant with the curving form of the data. This in turn increases the performance on classification with the model having a test accuracy of nearly 87%. The above improvement is an indication that the polynomial kernel can represent non-linear relationships that a linear classifier cannot render (Lu et al., 2021).

Despite these benefits, the use of poly-kernel introduces new hyperparameters, including the degree of the poly-kernel and the strength of regularisation which have a direct effect on the complexity of models. A model which is overly high could be overfitted to the training data and therefore have a lower likelihood of generalising to new samples. As a result, hyperparameters and tuning should be carefully considered when applying to the use of a polynomial kernel especially in the instance of a small or noisy data.

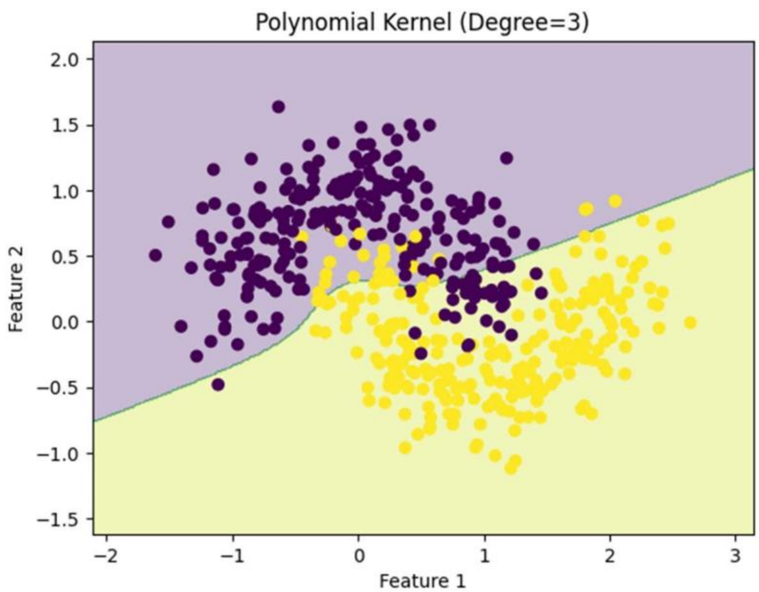

Figure 3: Decision boundary produced by a polynomial kernel SVM with degree 3

7. Radial Basis Function (RBF) Kernel

One of the most popular kernel functions in Support Vector Machines is the Radial Basis Function (RBF) kernel because it has a very high level of flexibility and empirical effectiveness when applied to a variety of classification problems. The RBF kernel implicitly projects input data to an infinite-dimensional feature space, and the model can learn very complicated and non- linear decision boundaries, but it does not actually transform the input data.

Using the two-moons data, the RBF kernel is shown to possess a significant superiority over the linear and the poly-nomial kernels. The decision boundary resulting is smooth but highly adaptive and the boundary to be followed closely follows the underlying curved structure of the data. It results in a significant enhancement of classification efficiency, where test accuracy is about 96 - 97%, the best-performing kernel in this paper (Deepa et al., 2023).

One of the most important hyperparameters added by the RBF kernel is gamma that regulates the impact of a single training example on the decision boundary. Higher gamma values bring out more complex boundaries that can closely fit the training data hence higher chances of overfitting. On the other hand, lower gamma values result in less jagged edges that will not be able to fit the data. The choice of a proper gamma is hence important to have a balance between the flexibility of the model, and the performance of generalisation.

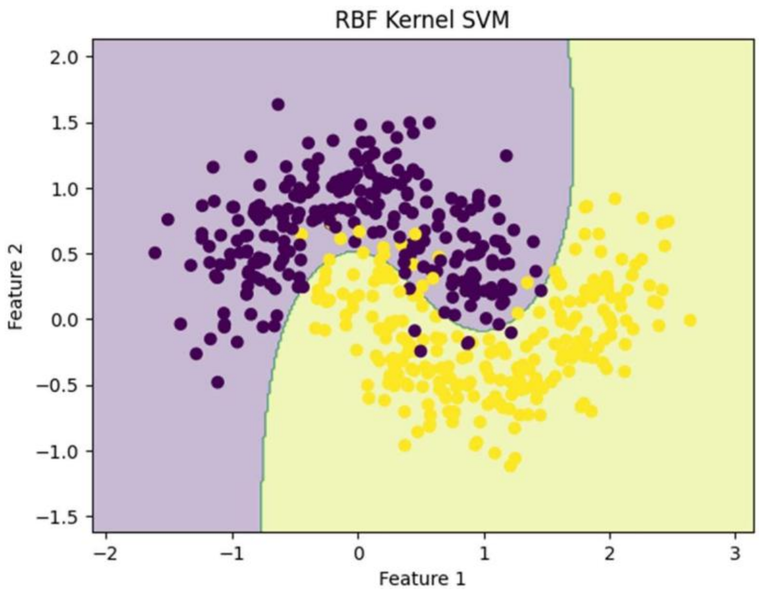

Figure 4: Decision boundary produced by an RBF kernel SVM

8. Hyperparameter Effects and Bias–Variance Trade-off

To explore the Support Vector Machine behaviour further, the regularisation parameter C was changed under the Radial Basis Function Radial Basis Function (RBF) kernel. The parameter C involves the trade-off of the aim to maximise the margin and reduce classification errors. Three values have been assessed: 0.1, 1 and 10.

A smaller C (= 0.1) puts more focus on a larger margin, and more misclassifications have been permitted, giving path to a decision boundary that is less jagged. This structure is more biased and has a lower test accuracy of about 91 which means under fitting (Lang & Lu, 2024). A higher parameter of C = 1 gives a reasonable bias vs variance trade-off, a more flexible trade-off boundary and much better performance with an accuracy of about 96. Increasing the value further to C = 10 further penalizes points misclassified greatly making the line more complex and results in a slight improvement in accuracy of approximately 97%.

These results are the strong evidence of the bias (variance) trade-off, which is one of the basic concepts of machine learning. Although more complex model complexity may improve its performance in training data, more complex model complexity also raises the probability of overfitting. This implies that it is necessary to hyperparameter tune so that it generalises well on unknown data.

9. Ethical and Practical Considerations

Although the Support Vector Machines are both strong and efficient classification models, their implementation creates a number of ethical and practical concerns that should be observed keenly. Some of the non-linear kernels utilized in kernel-based SVMs such as the Radial Basis Function (RBF) do not have an interpretable nature as linear models do. This lack of transparency may be dangerous in areas that require substantial stakes like health care, finance, and criminal justice, where it is necessary to comprehend and provide a description of model choices to hold these areas accountable and to gain trust.

The quality and representativeness of the training data is also very crucial in the performance of SVM. When the dataset inclusive of these biases or inability to capture all the groups in the population, the formulated model can perpetrate or even increase the existing inequalities in society. The use of machine learning must be conducted in a thoughtful way by ensuring the dataset is chosen, preprocessed, and validated after which the model fairness and possible discriminatory behavior should be constantly monitored (Yang et al., 2023).

Practically, it is possible that the computation of kernel-based techniques can be computationally demanding including when the method is used on large-scale data sets. Scalability is also a major issue as training time and memory required to reach large dataset sizes scale in a nonlinear fashion. Where resource constraints or real-time decision-making demands are involved, practitioners need to trade-off predictive performance and computational efficiency and look to other models.

10. Conclusion

This tutorial showed the effects of the various SVM kernels in decision boundary and classification with a non-linearly separable dataset. The linear kernel proved unsuitable to be able to underfit complex data, and the medium quality increase in quality in the polynomial kernel made it more searches more flexible. RBF kernel came up with the best performance since it was able to capture the non-linear relationship.

By applying visualisation and systematised experimentation, this tutorial showed that it is essential to select the kernels and optimise hyperparameters of SVMs. This knowledge of these factors will help practitioners make better modelling decisions and will help them better apply SVMs in their own work.

References

Shafi, I., Din, S., Khan, A., De La Torre Díez, I., Del Jesús Palí Casanova, R., Pifarre, K. T., & Ashraf, I. (2022). An Effective Method for Lung Cancer Diagnosis from CT Scan Using Deep Learning-Based Support Vector Network. *Cancers*, *14*(21), 5457. https://doi.org/10.3390/cancers14215457

Kok, Z. H., Shariff, A. R. M., Alfatni, M. S. M., & Khairunniza-Bejo, S. (2021). Support Vector Machine in Precision Agriculture: A review. *Computers and Electronics in Agriculture*, *191*, 106546. https://doi.org/10.1016/j.compag.2021.106546

Anand, P., Deb, C., Yan, K., Yang, J., Cheong, D., & Sekhar, C. (2021). Occupancy-based energy consumption modelling using machine learning algorithms for institutional buildings. *Energy and Buildings*, *252*, 111478. https://doi.org/10.1016/j.enbuild.2021.111478

Su, M., Peng, H., & Li, S. (2021). A visualized bibliometric analysis of mapping research trends of machine learning in engineering (MLE). *Expert Systems With Applications*, *186*, 115728. https://doi.org/10.1016/j.eswa.2021.115728

Mastouri, A., Zhu, Y., Gultchin, L., Korba, A., Silva, R., Kusner, M., Gretton, A., & Muandet, K. (2021, July 1). *Proximal Causal Learning with Kernels: Two-Stage Estimation and Moment Restriction*. PMLR. https://proceedings.mlr.press/v139/mastouri21a.html

Lu, F., Maggioni, M., & Tang, S. (2021). *Learning interaction kernels in heterogeneous systems of agents from multiple trajectories*. https://www.jmlr.org/papers/v22/19-861.html

Deepa, S. N., Natarajan, N., & Berlin, M. (2023). Enhanced variational mode decomposition with deep learning SVM kernels for river streamflow forecasting. *Environmental Earth Sciences*, *82*(22). https://doi.org/10.1007/s12665-023-11222-5

Xu, N., Liu, F., & Sutherland, D. J. (2024, September 10). *Learning representations for independence testing*. arXiv.org. https://arxiv.org/abs/2409.06890

Lang, Q., & Lu, J. (2024, February 18). *Learning memory kernels in generalized Langevin equations*. arXiv.org. https://arxiv.org/abs/2402.11705

Yang, S., Guo, N., Santha, M., & Rebentrost, P. (2023). Quantum Alphatron: quantum advantage for learning with kernels and noise. *Quantum*, *7*, 1174. https://doi.org/10.22331/q-2023-11-08-1174