|  |  |
| --- | --- |
|  |  |
|  | Machine Learning |
|  |  |
|  |  |
|  | **Lab 2: Support Vector Machines**  LinusGroß  DanielMensah |
|  |  |

|  |
| --- |
|  |

# Assignment 1

*Move the clusters around and change their sizes to make it easier or harder for the classifier to find a decent boundary. Pay attention to when the optimizer (minimize function) is not able to find a solution at all.*

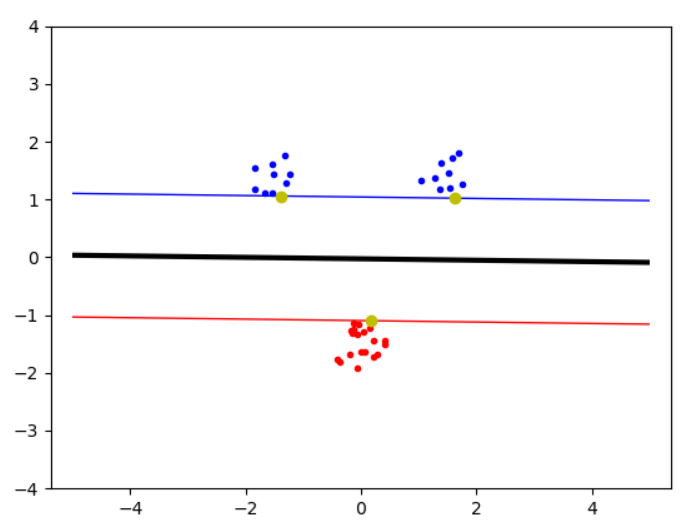


Figure 1: Easy classification because there is a clear boundary between the datasets. Furthermore

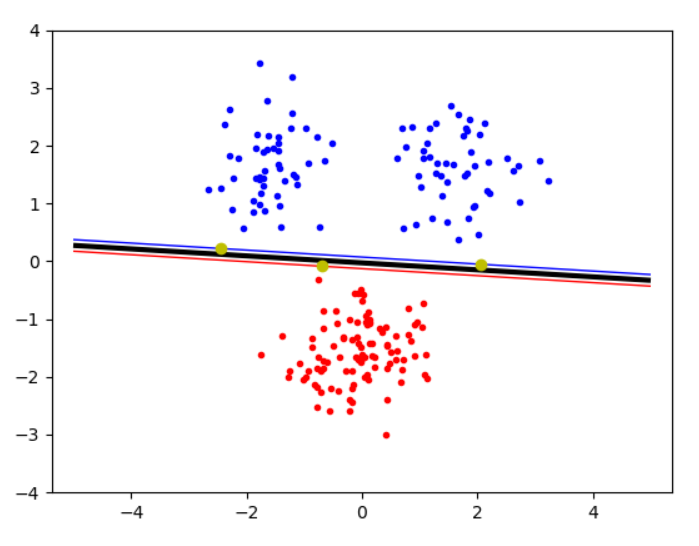


Figure 2: High variance and large clusters -> very small margin, because without slack the widest datapoints become the support vectors

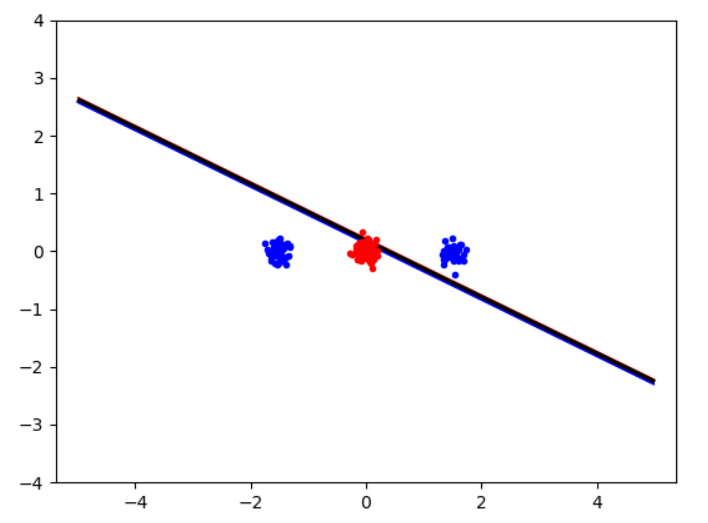


Figure 3: Data is not linear separable, the algorithm finds no solution

# Assignment 2

*Implement the two non-linear kernels. You should be able to classify very hard data sets with these.*

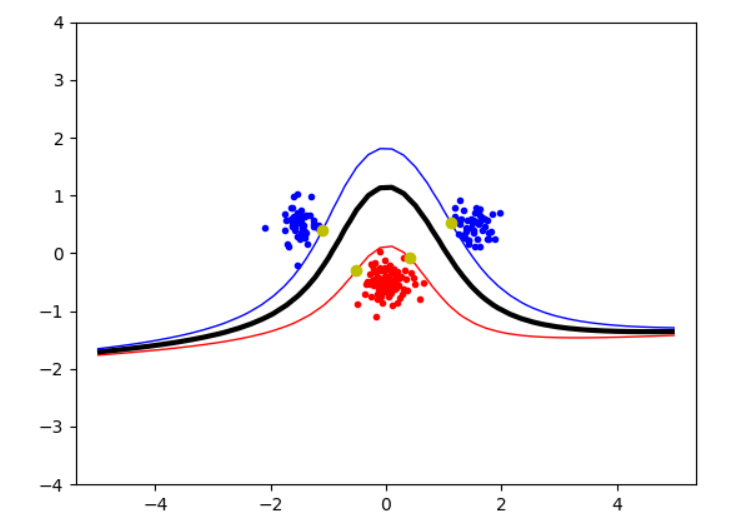


Figure 4: Polynomial kernel, p = 2

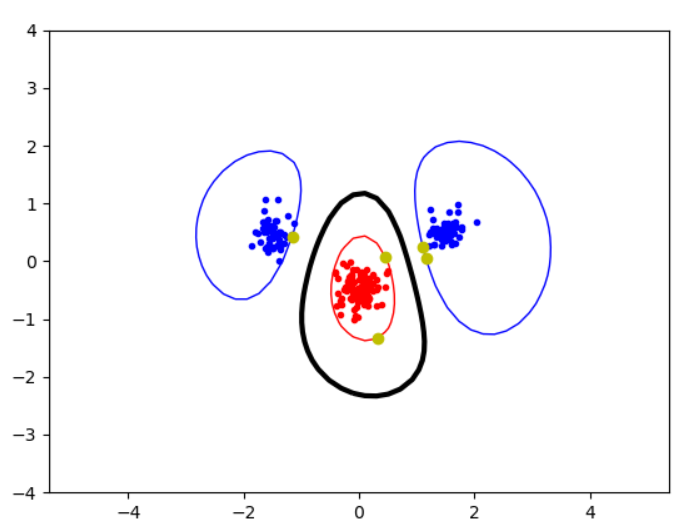


Figure : RBF kernel, = 1

Using the more complex kernels, the decision boundaries are not forced to be linear. This leads to curvy decision boundaries, which can classify more complex datasets, where there was no linear separation possible.

# Assignment 3

*The non-linear kernels have parameters; explore how they influence the decision boundary. Reason about this in terms of the bias-variance trade-off.*

Polynomial kernel:

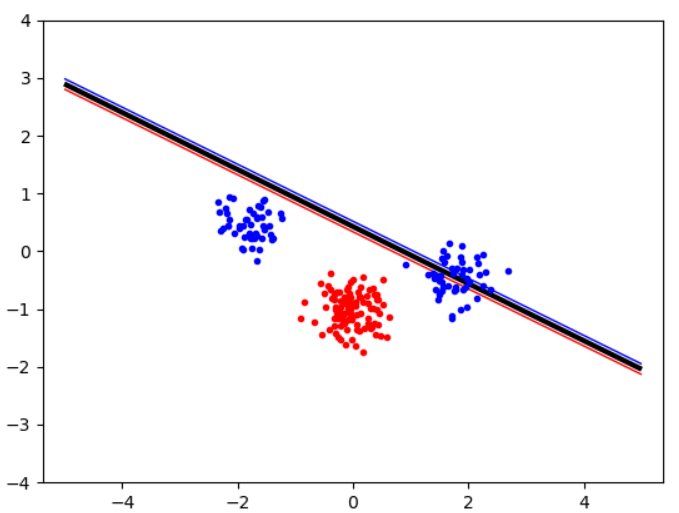


Figure 6: polynomial kernel, p = 1: linear kernel, no solution is found

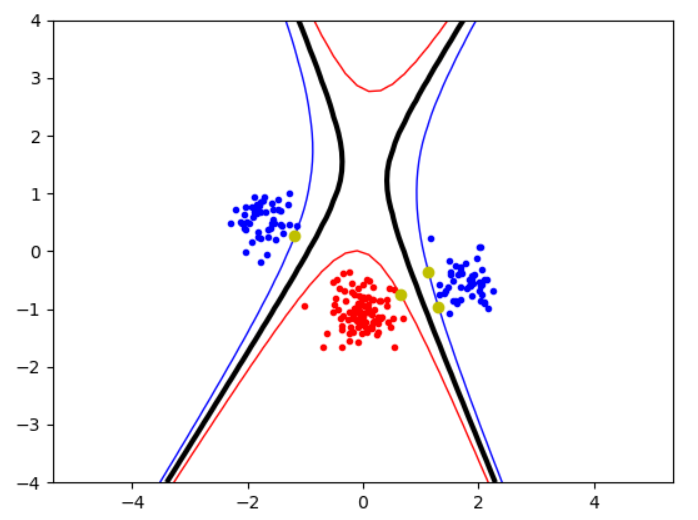


Figure 7: polynomial kernel, p = 2: no satisfying boundary

Radial Basis Function (RBF) kernel:

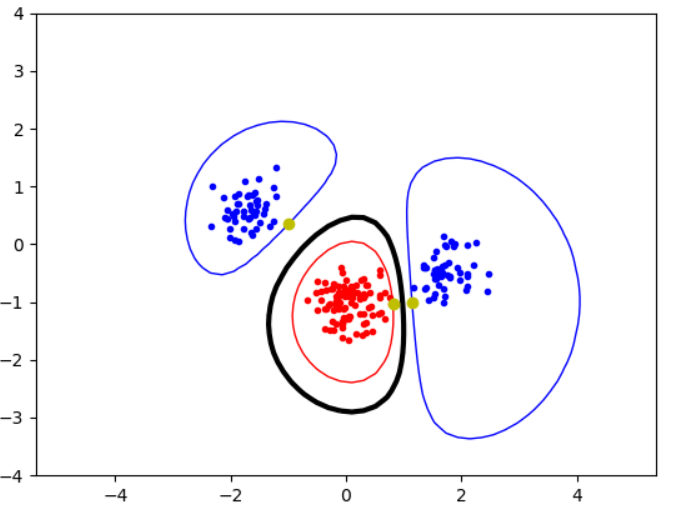


Figure 8: RBF-kernel, σ = 1: very good classification with round decision boundaries around the clusters

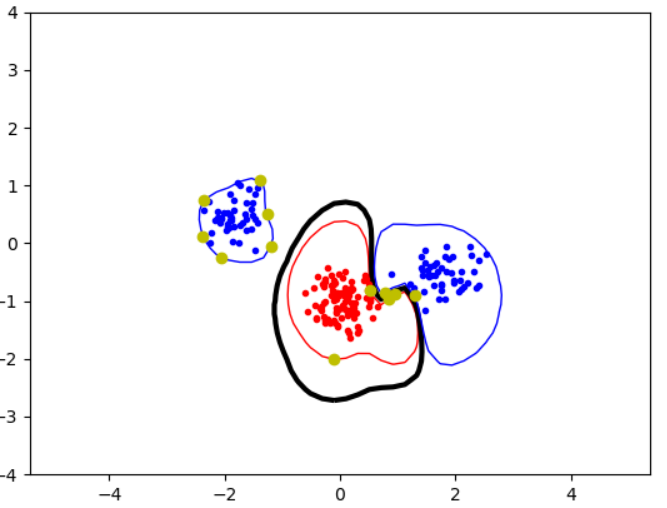


Figure 9: RBF-kernel, σ = 0.5: the decision boundaries get less smooth and more support-vectors are used

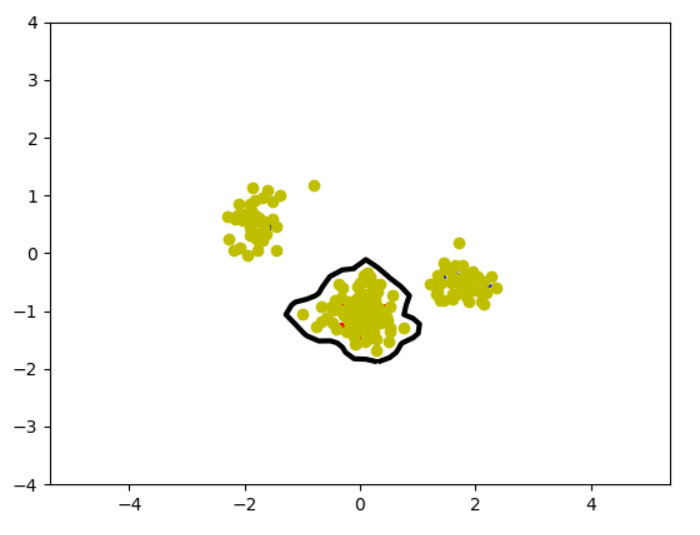


Figure 10: RBF-kernel, σ = 0.1: almost all datapoints are used for the decision boundary

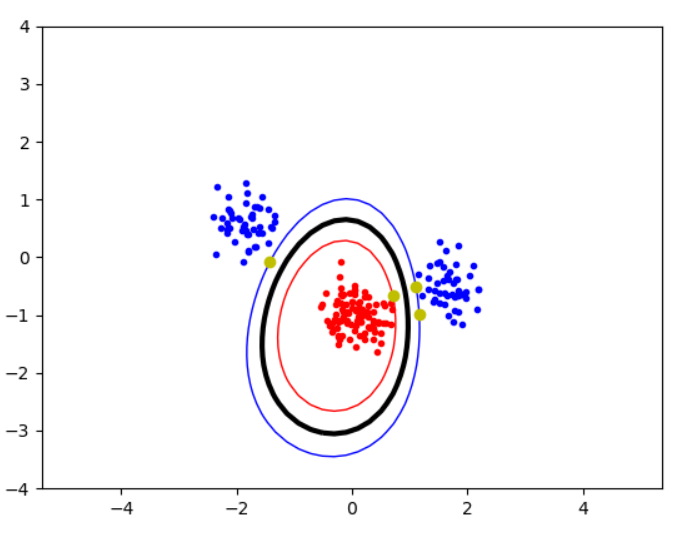


Figure 11: RBF-kernel, σ = 2: more clear decision boundaries which allow for a higher variance

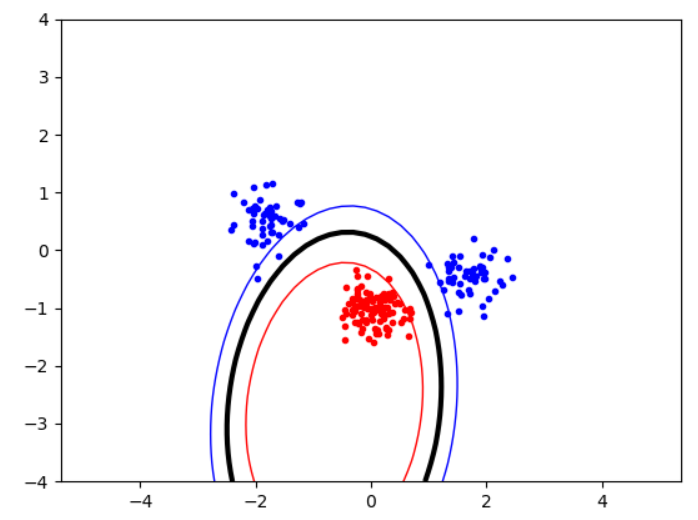


Figure 12: RBF-kernel, σ = 10: the optimizer is not able to find a solution without slack

For the RBF-kernel, the parameter σ determines the variance-bias tradeoff.

If you decrease σ, the decision boundary takes more datapoints into count which leads to smaller decision boundaries. The bias increases, because there is a clearer classification, the variance however decreases, because the classes are not allowed to be widely spread.

Increasing σ leads to the opposite effect: The decision boundaries get wider and allow for a higher variance. Correspondingly the bias decreases. However, if the parameter is raised to high, the optimizer is not able to find a solution, if no slack is allowed. This is because the allowed variance is to big and would lead to datapoints of the training set within the margin.

# Assignment 4

*Explore the role of the slack parameter C. What happens for very large/small values?*

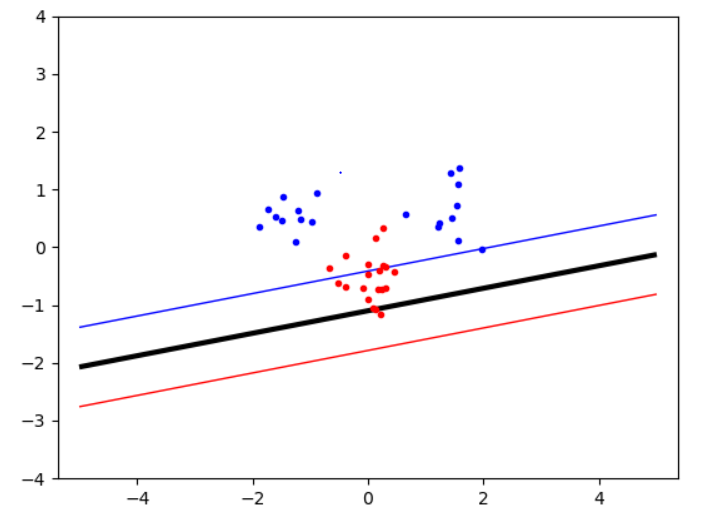
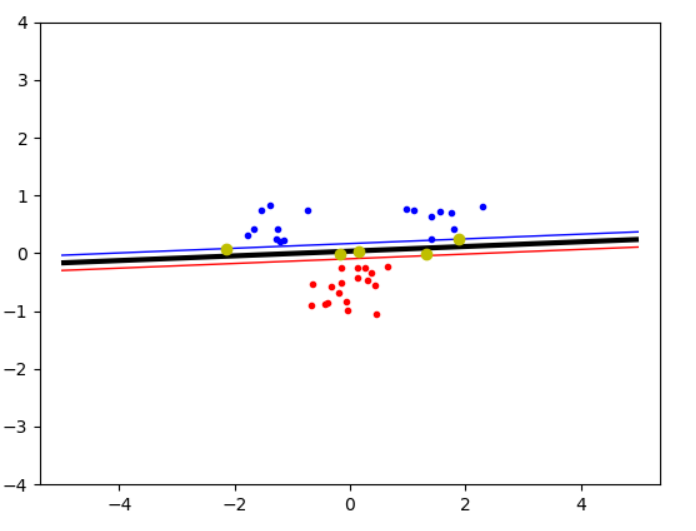


Figure 13: C = None, without the slack-parameter, no linear solution can be found



# Assignment 5

*Imagine that you are given data that is not easily separable. When should you opt for more slack rather than going for a more complex model (kernel) and vice versa?*