DD2421
Machine Learning

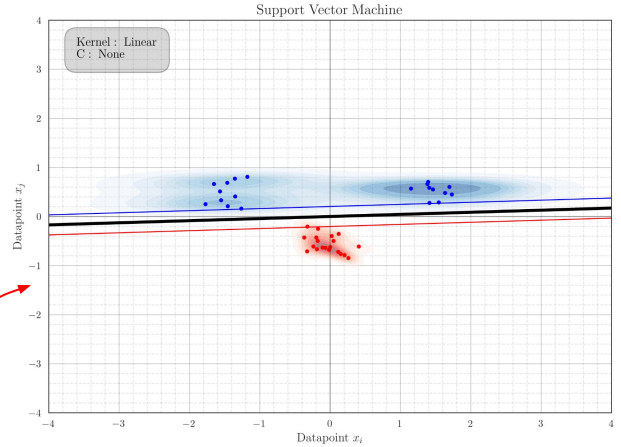
Support Vector Machines

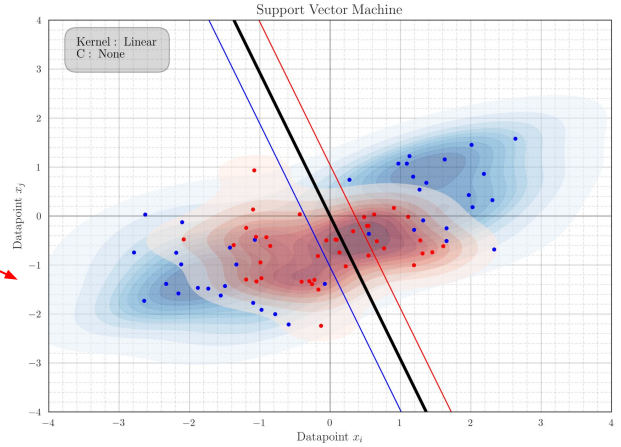
1. Move the clusters around and change their sizes to make it easier or harder for the classifier to find a decent boundary. Pay attention to when the optimizer (`minimize` function) is not able to find a solution at all.

Clusters that are linearly separable causes the optimizer to find a solution for the clusters.

As we introduce non-linear kernels, the optimizer is again able to find a solution for the decision boundary.

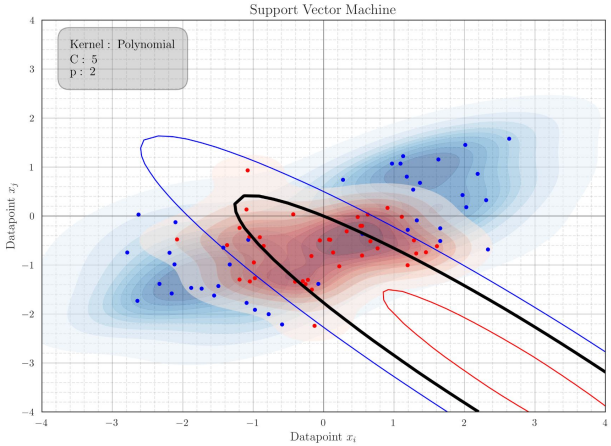
Successful optimization
vs
Unsuccessful optimization



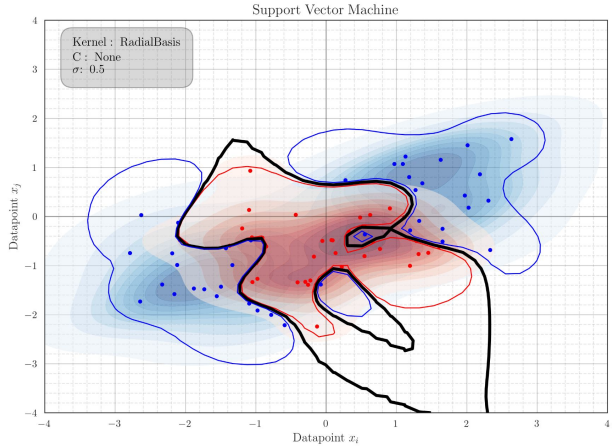


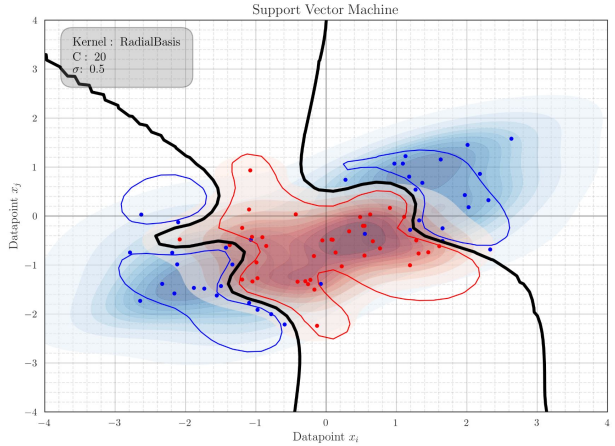
2. Implement the two non-linear kernels. You should be able to classify very hard data sets with these.

Polynomial Kernel



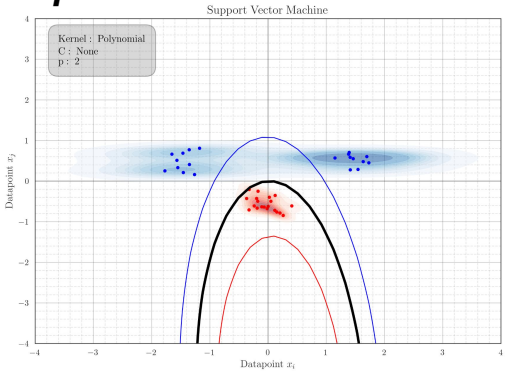
Radial Basis Kernel



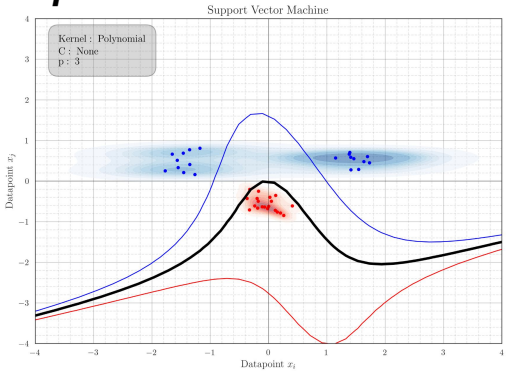


3. The non-linear kernels have parameters; explore how they influence the decision boundary. Reason about this in terms of the bias-variance trade-off.

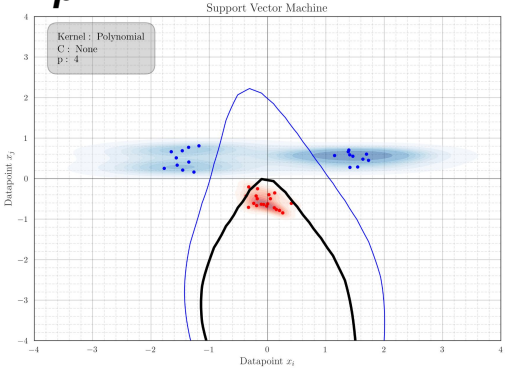
***p* = 2**



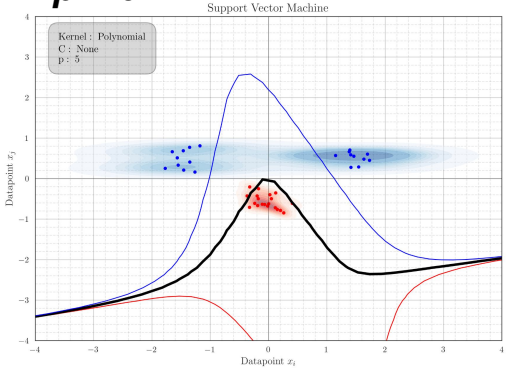
***p* = 3**



***p* = 4**



***p* = 5**



Polynomial Kernel
Changing the order (degree) of the polynomial by changing the power *p*.

p = 2: quadratic shapes (ellipses, parabolas, hyperbolas)
p > 2: more complex shapes.

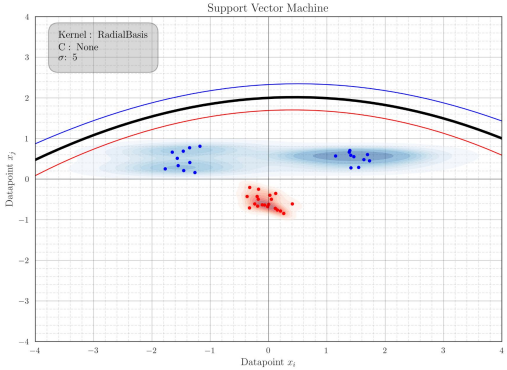
As *p* increases the bias**(?)** decreases and variance**(?)** increases.

A higher *p* makes a more complex model.

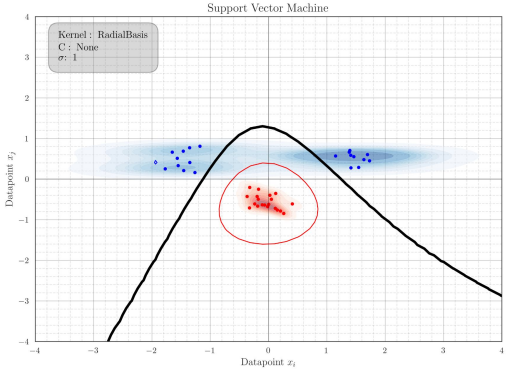
$$\mathcal{K}(\vec{x}, \vec{y}) = (\vec{x}^T \cdot \vec{y} + 1)^p$$

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.

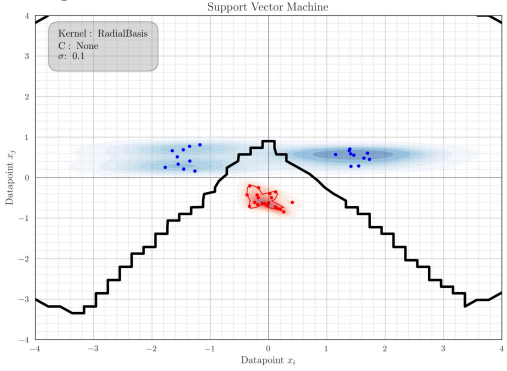
σ = 5



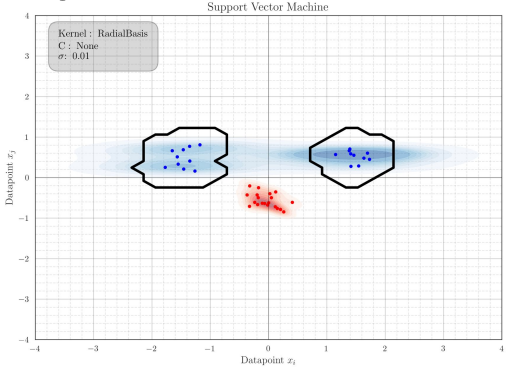
σ = 1



σ = 0.1



σ = 0.01



Radial Basis Kernel

The parameter **σ** is used to control the smoothness of the boundary.

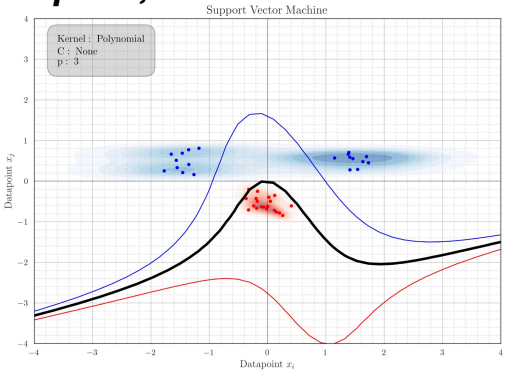
As **σ** decreases the bias decreases and the variance increases.

A smaller **σ** makes a more complex model.

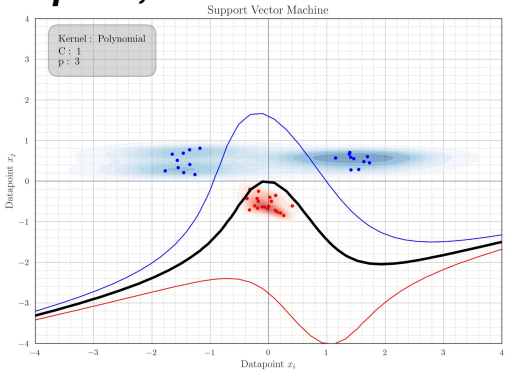
$$\mathcal{K}(\vec{x}, \vec{y}) = e^{-\frac{||\vec{x}-\vec{y}||^2}{2\sigma^2}}$$

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

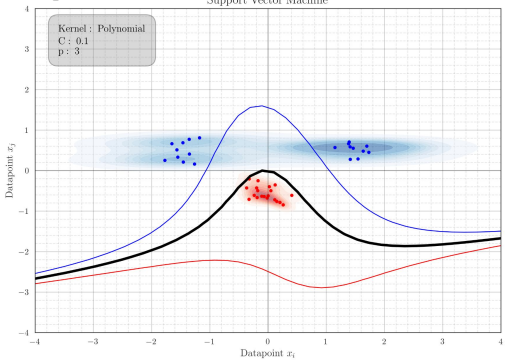
p = 3, C = None



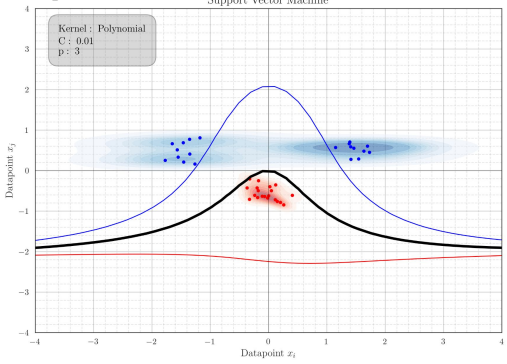
p = 3, C = 1



p = 3, C = 0.1



p = 3, C = 0.001



Polynomial Kernel & Slack

By controlling the slack, the margin is controlled, by limiting the maximum values of α.

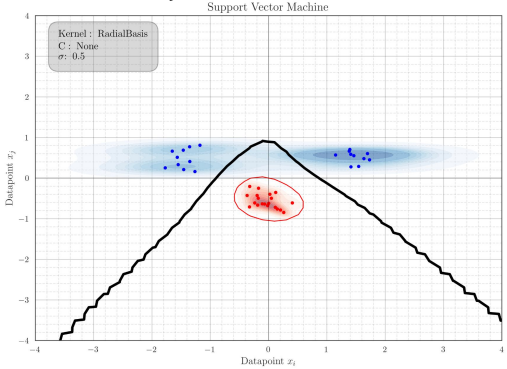
Decreasing the *C* when using the *polynomial kernel*
→ *increases the slack*
→ decreases the complexity of the model, and makes the decision boundary smoother.

By making the complex less complex, the variance is decreased and bias increased.

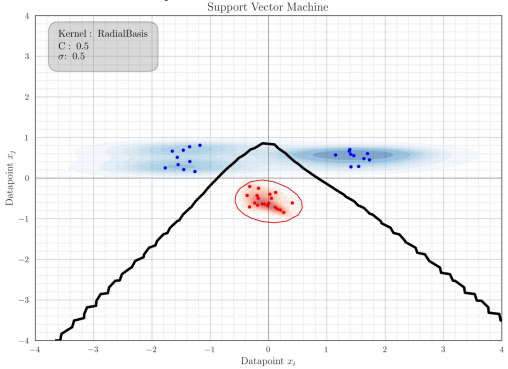
Large values of slack has no impact on the model, since it does not limit any used α.

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

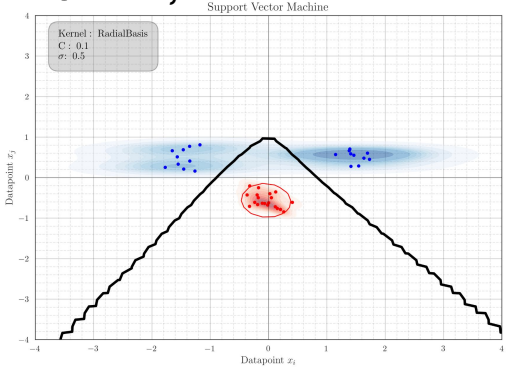
σ = 0.5, C = None



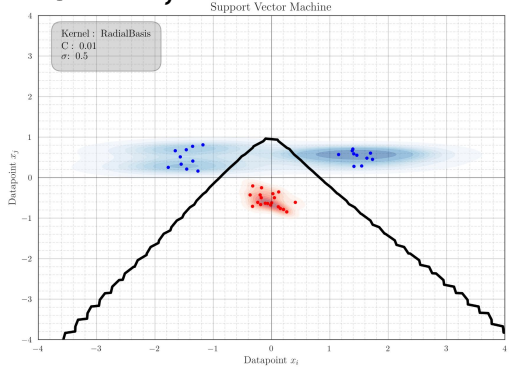
σ = 0.5, C = 0.5



σ = 0.5, C = 0.1



σ = 0.5, C = 0.01



Radial Basis Kernel & Slack

Decreasing *C* when using the *radial basis kernel*

→ increases the slack

→ increases the margin by decreasing the maximum values of α.

This causes the model to be less complex. Therefore, the variance is decreased and the bias is increased.

Next page for more clear differences

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

Radial Basis Kernel & Slack

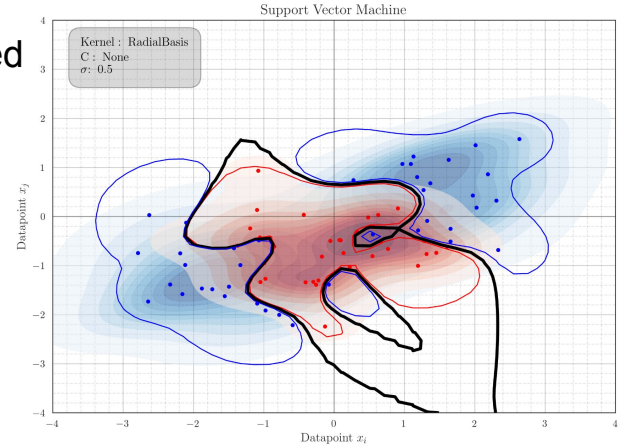
Decreasing *C* when using the *radial basis kernel*
→ increases the slack
→ increases the margin by decreasing the maximum values of α.

This causes the model to be less complex. Therefore, the variance is decreased and the bias is increased.

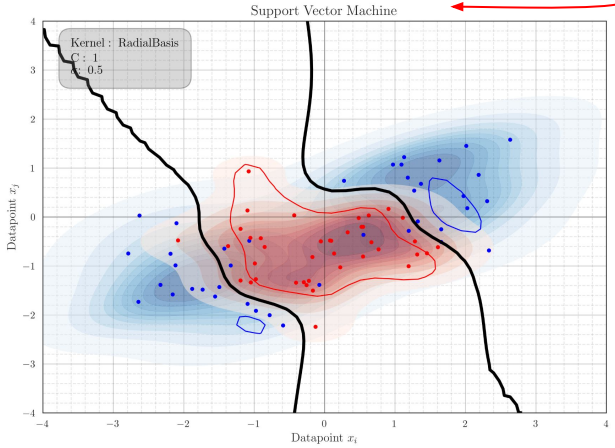
Specialized

Generalized

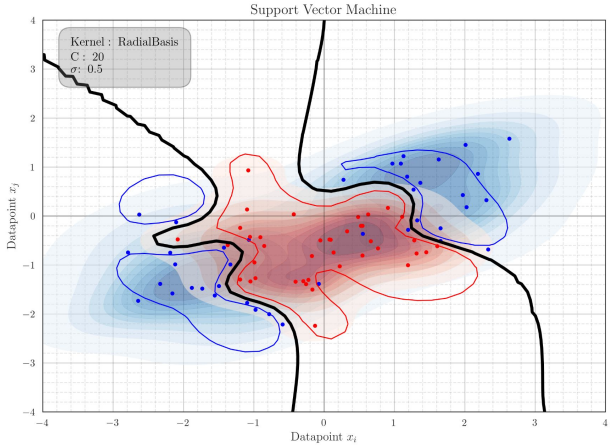
σ = 0.5, C = None



σ = 0.5, C = 1



σ = 0.5, C = 20



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?

When the center of the datapoint classes are close, and noise are present, introducing more slack is beneficial to avoid misclassification of true points.

This might cause datapoints far off the center to be misclassified, but this is a trade-off that we have to accept.