# Machine Learning - Lab 2 (Support Vector Machines)

hfag@kth.se

March 9, 2022

# 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.**

The linear Kernel is used,

$$K(\vec{x}, \vec{y}) = \vec{x}^T * \vec{y}.$$

Default settings, i.e. the ones that was given in the lab description (1 apart in height):

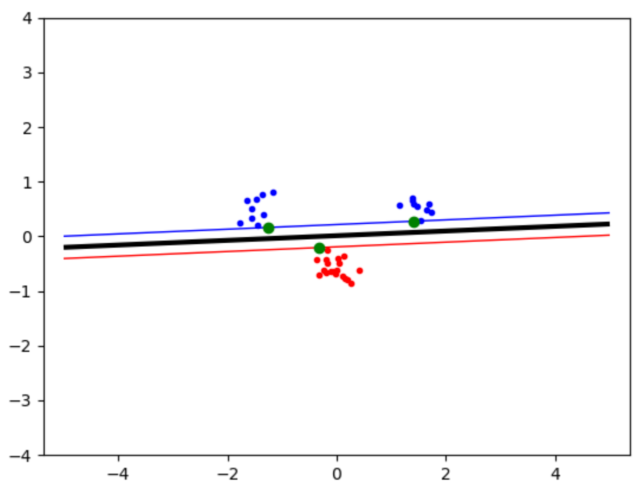


Figure 1: Linear with ”default” settings.

Move class A, 1 unit up (2 apart in height):

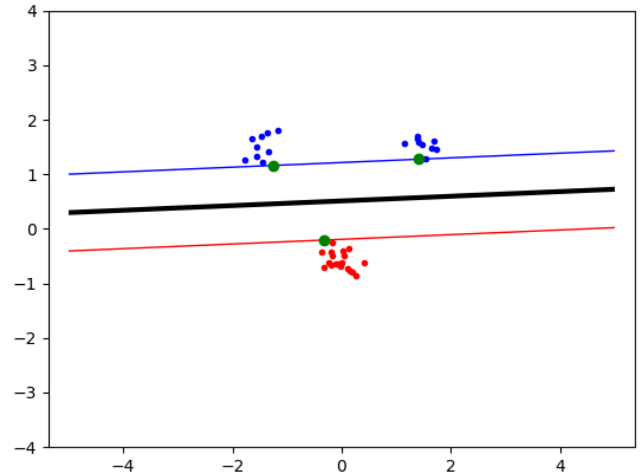


Figure 2: Linear, moved class A cluster away from class B.

Move class A, 0.5 unit down (0.5 apart in height):

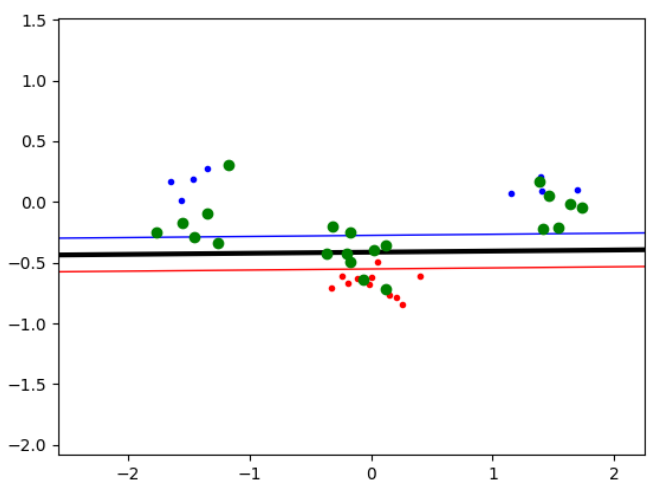


Figure 3: Linear, moved class A cluster towards class B.

Increase size for class A by 0.2:

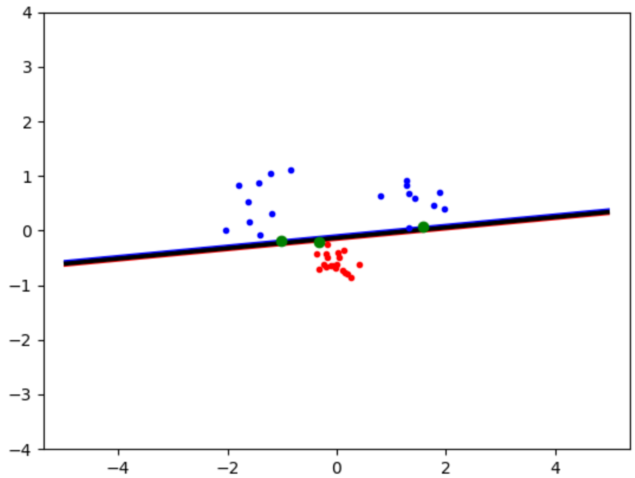


Figure 4: Linear, increased class A cluster size-term with 0.2.

Increase size for class B by 0.2:

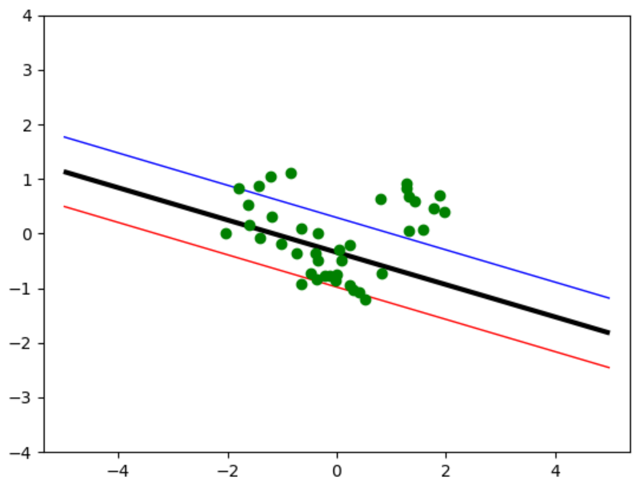


Figure 5: Linear, increased class B cluster size-term with 0.2.

# Assignment 2

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

$$K(\vec{x}, \vec{y}) = \left(\vec{x}^T * \vec{y} + 1\right)^P \tag{1}$$

$P = 1$ - Linear
$P = 2$ - Quadratic shapes
$P = 3$ - "Complex" shapes

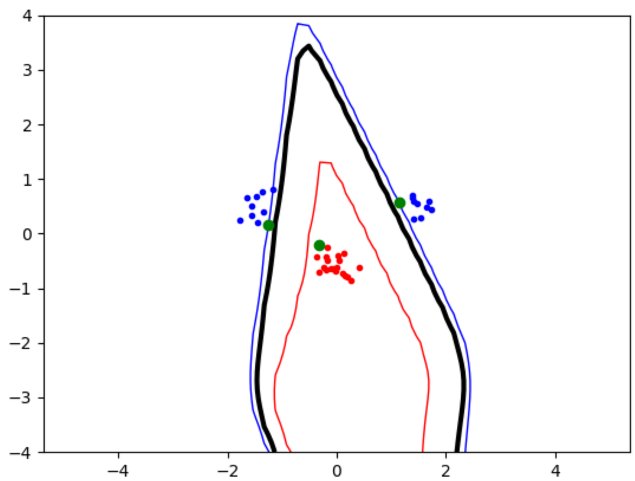


Figure 6: Polynomial Kernel with $P = 10$ (default settings).

Radial Basis Function kernel. This uses euclidian distance between the two datapoints, with parameter $\sigma$ to control smoothness of the boundary.

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

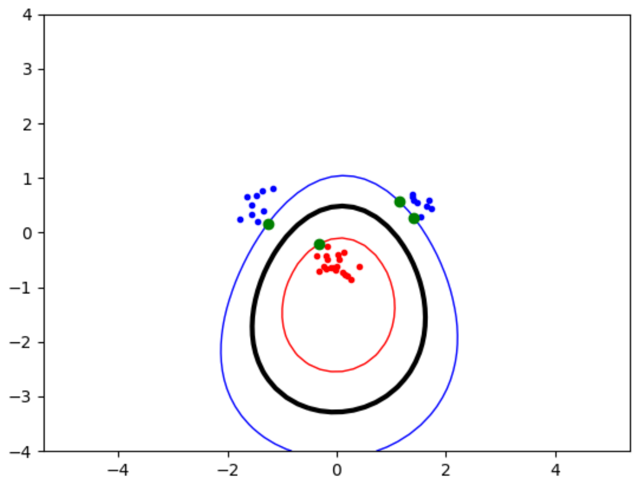


Figure 7: RBF Kernel with $\sigma = 2$(default settings).

## 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

As we increase the polynomial degree, the risk of overfitting the data becomes larger. E.g. for a dataset with a lot of points, perhaps a 15 degree polynomial will be well-fitted as it increases the margin the best but looses a great amount of generality since the problem most likely is not a 15 degree polynomial. Thus as we increase the p we increase the variance and make it more sensitive to small fluctuations and therefore decrease the bias.

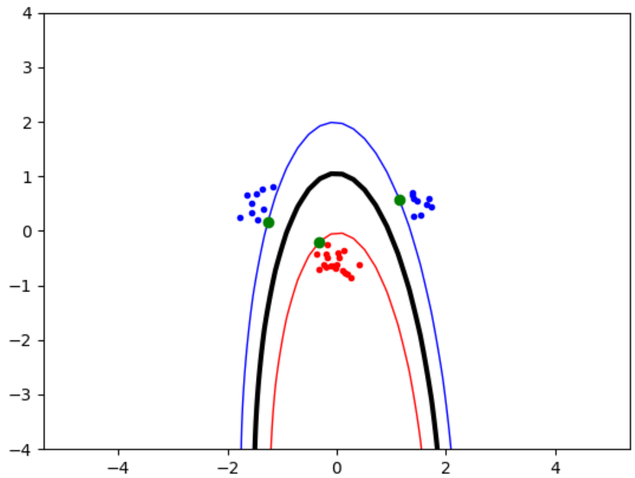


Figure 8: Polynomial Kernel with $P = 2$ (default settings).

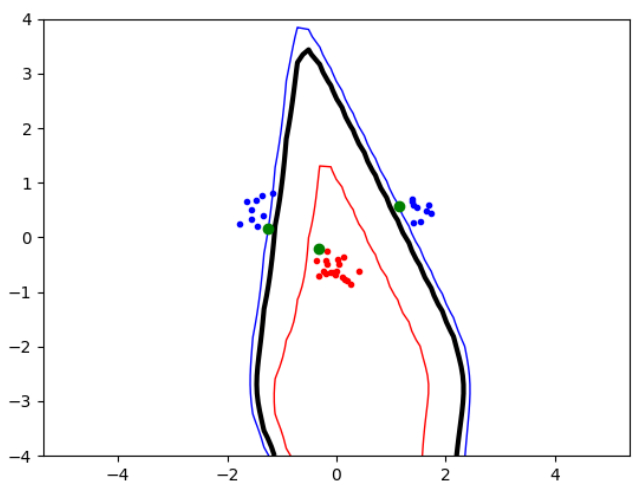


Figure 9: Polynomial Kernel with $P = 10$ (default settings).

## RBF

First of all we note if sigma is low the bias will be low and variance high, and vice versa for high sigma. Since, given a small sigma the Kernel is sensitive to small fluctuations (overfitting), whilst if we have a high sigma the Kernel have a greater generality (underfitting).

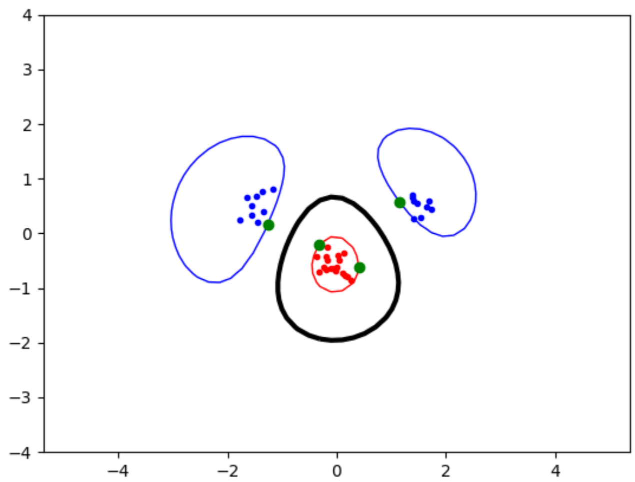


Figure 10: RBF Kernel with $\sigma = 1$ (default settings).

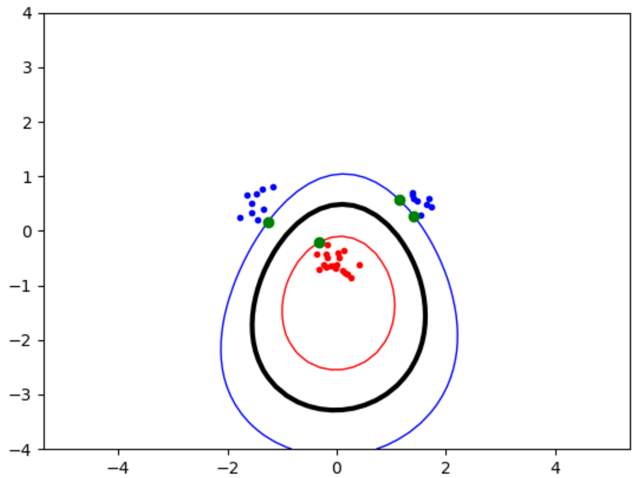


Figure 11: RBF Kernel with $\sigma = 2$ (default settings).

# Assignment 4

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

The slack variable allows for misclassification in the algorithm in order to maintain some ”generality” in the model. A difficult problem, i.e. where there exist a lot of noise in the data, requires a slack variable for a boundary to be found. The following figures show that the model is approaching ideal if there were no noise for higher values of C.

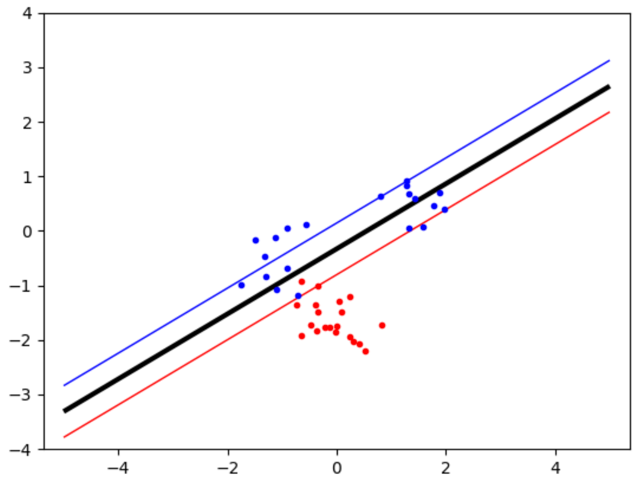


Figure 12: $C = None$.

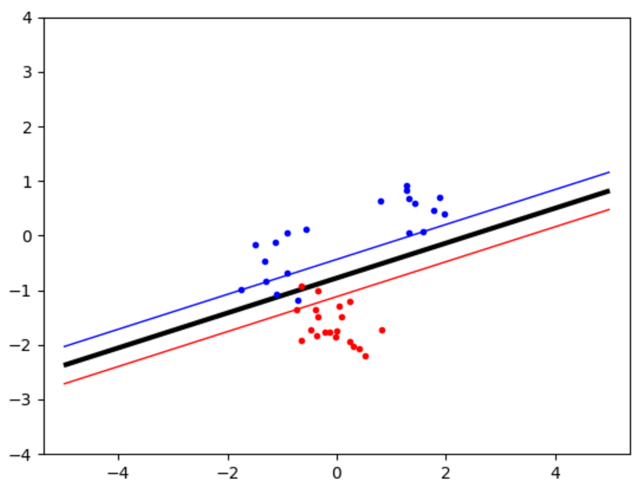


Figure 13: $C = 10$.

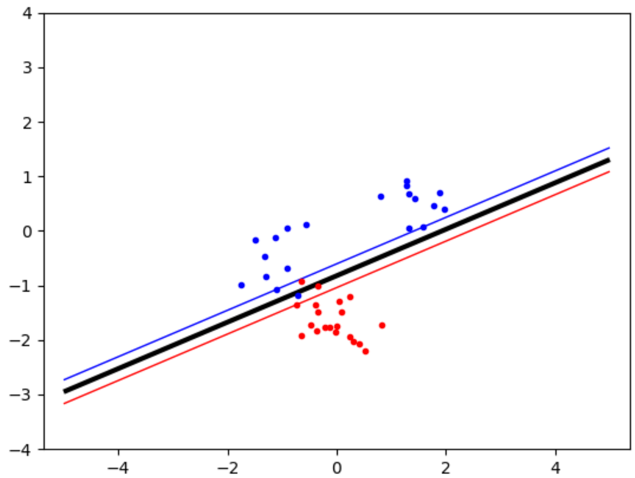


Figure 14: $C = 100$.

# 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?**

If there can be seen some kind of linear tendencies but possibly have a lot of noise, a less complex model with slack is preferred. If there clearly is no linear relationship can be seen, then a more complex model should be used.

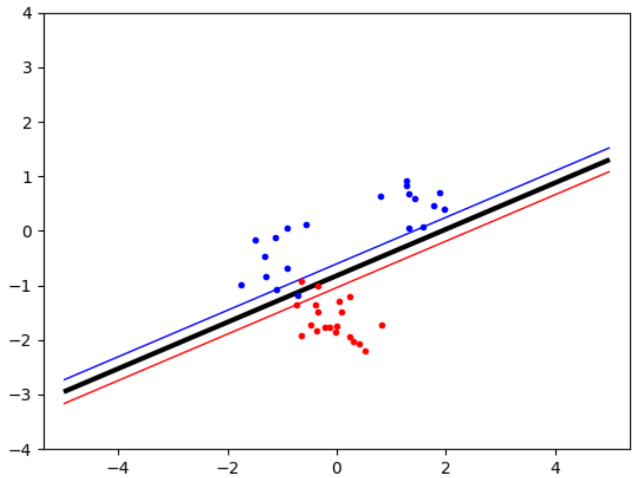


Figure 15: Linear tendencies, slack should be used.

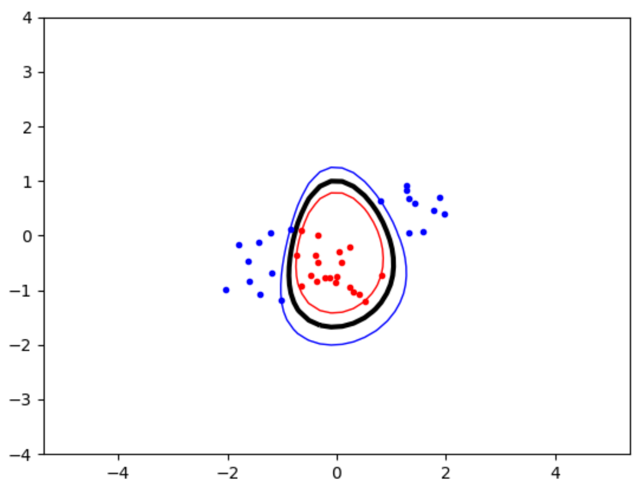


Figure 16: No linear tendencies, complex model (RBF) should be used.