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

We didn't find any solution when the svm couldn't find a line that separated the classes. We increased the variance from 0.2 to 0.4, changed the size and increased the cluster further away in order to find a line that separated the classifications.

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 .

For the polynomial function when p=2 and p=3, we got the best result. For the radial bias function we got a good separator at 0.8.

By using the right parameter for the function we can make the svm more accurate.

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

Since some training data usually has some misclassification or noise, the purpose for C is to avoid overfitting.

When C is high it gives a low bias and a high variance, overfitting. For small C it gives a high bias and low variance, underfitting.

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?

You should opt for more slack when you have a complex model, since higher slack will give a low bias and high variance(overfitting).

When you have a model with high slack you should opt for a more complex model, since the model might be too simple.