**Group 1:**

* Xue Liu
* Soukaïna Mahboub Mehboub
* Dogukan Bagci
* Patrick Linnanen
* Goksu Sankal

**Assignment 2: Practice**

For this Assignment we did 5 Experiments, which can be divided in two parts. In the first part we compare the training error of SVM with a linear, polynomial (degree = 5) and gaussian kernel. For all the models c equals to 1. In the second part we compare the training error of a linear and a polynomial (degree = 5) kernel with c being equal to 1000. We used the [breast cancer](https://scikit-learn.org/stable/datasets/toy_dataset.html#breast-cancer-wisconsin-diagnostic-dataset) from sklearn. It has 569 instances and 30 features . In order to fulfill the criteria of this assignment we only used the first 5 features. Using the sklearn.decomposition.PCA class we reduced our feature matrix to only two features. The loss was calculated by using the sklearn.metrics.hinge\_loss method.

Part one:

| Kernel | C | Loss |
| --- | --- | --- |
| Linear | 1 | 0.4662 |
| Polynomial | 1 | 0.5360 |
| Gaussian | 1 | 0.4569 |

These were the results of the first part. As you can see the Gaussian Kernel performed best by a slight margin. The Polynomial Kernel performed worst by a relatively big margin. The reason might be because this is a binary classification problem. You can also see in the plot that the data points are linearly separable, so a polynomial kernel might overfit to the data.

Part two:

| Kernel | C | Loss |
| --- | --- | --- |
| Linear | 1000 | 0.4587 |
| Polynomial | 1000 | 0.4956 |

This table represents the results of the second part. The polynomial Kernel is again the worst choice for this dataset. Both the linear and polynomial improved compared to the first part. This is because of the higher value of c, which causes that the hypotheses for both versions adapt more to the data.

The results and the code for this part can be found in the SVM\_Assignment\_Group1.ipynb file.