Task 1:

Q1: a) To be able to classify, Support Vector Machines separate *n*-dimensional data points depending on their classes using a *n*-1-dimensional hyperplane. Since there are an infinite number of possible hyperplanes between the separation, the best choice will be the one that represents the largest separation between the nearest data points on either side of the hyperplane. This separation between the classes is called the margin.

b) Support Vectors are those data points that are closer to the hyperplane or on the margin.

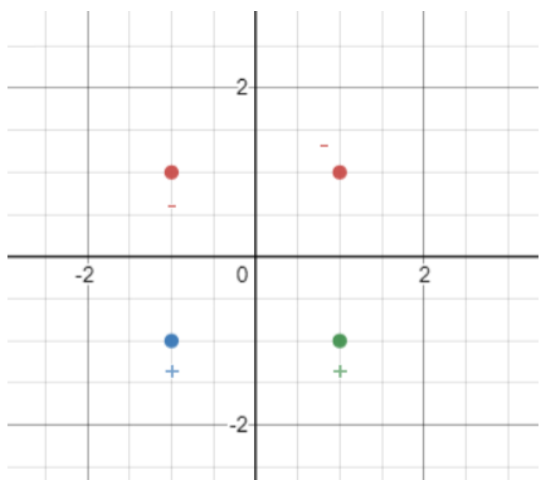
Q2-Q3: Depending on the dataset, there are two possible solutions for non-separable data. a) Soft margin, which lets some datapoints cross their decision boundaries and the SVM can tolerate this crossing. b) The Kernel trick which applies transformations to the dataset and allows the data to be projected in a higher dimensional space and consequently the data potentially becomes separatable. It is very common to use a combination of the two techniques.

Q4: When the kernel is applied to the data, this become a feature vector representation that can operate in a high-dimensional feature space.

Task 2:

First, we compute the values for x1.x2 and plot the figure.

x1	x2	x1.x2
-1	-1	1
-1	1	-1
1	-1	-1
1	1	1



Clearly, we can see now that the classes are linearly separable by the line x1.x2 = 0 with a margin of 1

Task 3:

Expansion:

$$x_1{}^2 + x_2{}^2 - 2ax_1 - 2bx_2 + (a^2 + b^2 - r^2) = 0$$

The weights correspond to the values 2a, 2b, 1, 1 which intercept in $a^2 + b^2 - r^2$. This shows that this kind of boundary would be linear separable in this feature space.

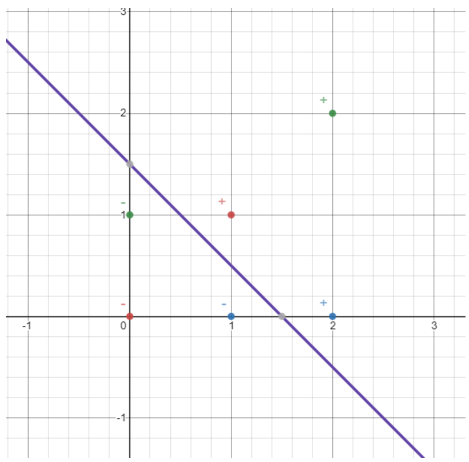
Task 4:

Expansion:

$$c{x_1}^2 + d{x_2}^2 - 2acx_1 - 2bdx_2 + (a^2c + b^2d - 1) = 0$$

The weights correspond to the values 2ac, 2bd, c, d, 0 which intercept in $a^2c + b^2d - 1$. This shows that this kind of boundary would be linear separable in this feature space.

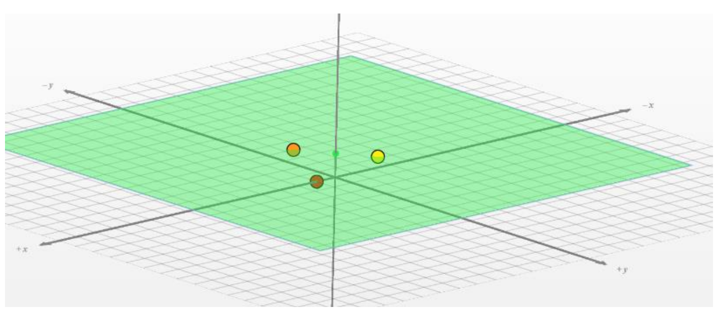
a) As seen in the following images, the classes are separable by the maximum hyperplane formed by the line x2 = 3/2 – x1



b) The equation of the hyperplane found is x2 = -x1 + 3/2 which has a slope of -1 and satisfies the values x1 = 3/2 and x2 = 0. Therefore, the weight vector is $(1, 1)^T$

Task 5:

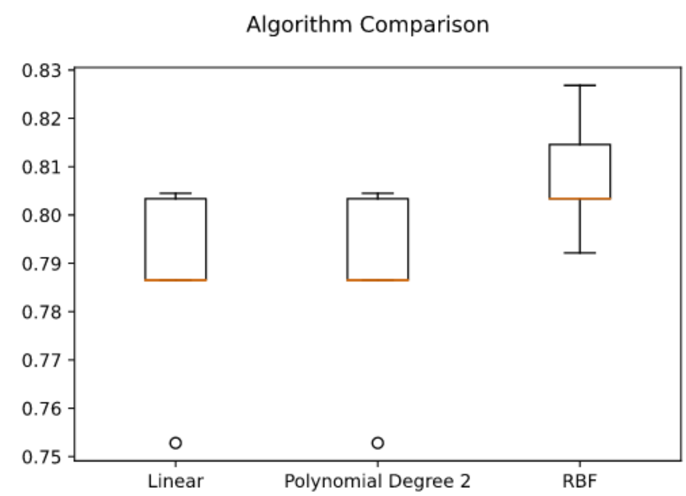
a) The classes are not separable in 1 dimension.
b) Now, the new points are (1,0,0), (1,-sqrt(2),1), (1, sqrt(2), 1) and a hyperplane could be placed in the point (0,0,1) as shown in the following image.



Task 6:

As we can see in the results below, linear kernel and quadratic kernel are more accurate than the RGB Kernel with previously unseen data. However, the cross-validation score is better in RGB. The linear and quadratic models might not generalize as well as the RGB model with a different set of data since they might have overfitted to that group of samples.

```
linear classification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       266
           1       1.00      1.00      1.00       152

    accuracy                           1.00       418
   macro avg       1.00      1.00      1.00       418
weighted avg       1.00      1.00      1.00       418

poly classification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       266
           1       1.00      1.00      1.00       152

    accuracy                           1.00       418
   macro avg       1.00      1.00      1.00       418
weighted avg       1.00      1.00      1.00       418

rbf classification report:
              precision    recall  f1-score   support

           0       0.94      1.00      0.97       266
           1       1.00      0.89      0.94       152

    accuracy                           0.96       418
   macro avg       0.97      0.95      0.96       418
weighted avg       0.96      0.96      0.96       418
```



For code, please visit https://github.com/lumalav/CAP5610/blob/master/HW4/HW4.pdf