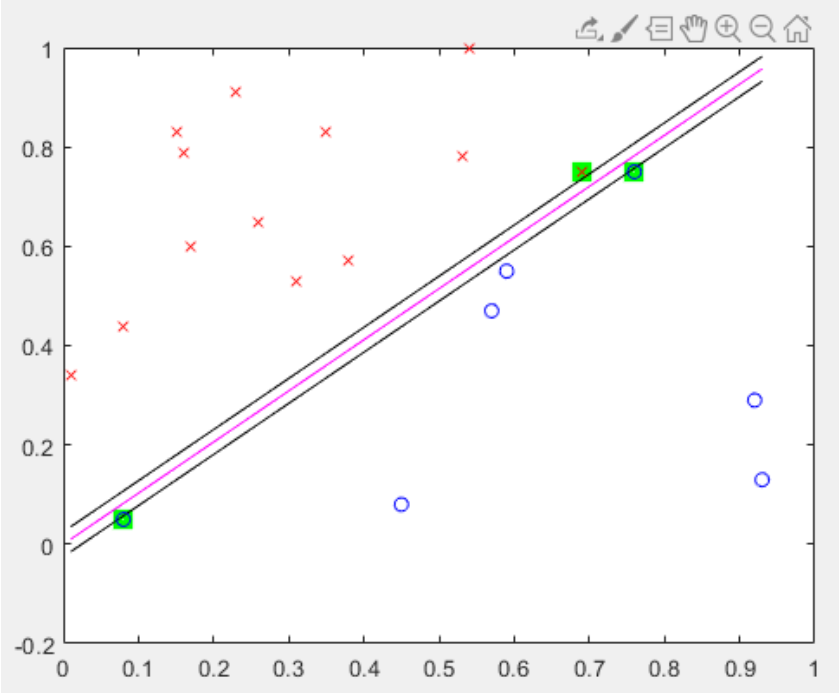
Introduction:

I think this assignment is about finding the most suitable decision boundary so that it is not too close to any of the points on either side. The distance from the closest point on both sides are equal in distance so that any new point added will be accurately marked as 1 or -1. Therefore, with this in mind, the support vector machine algorithm allows for better accuracy when finding a decision boundary that best fits all of the data.

Method:

The method used in this assignment was based on the quadprog() formula from Matlab. N is the number of total elements, d is the dimension size, and X is the N x d dataset. This formula required four parameters. First was H, which was a (d+1) x (d+1) matrix where the identity matrix was found inside this matrix starting at indexes (2,2). The second parameter, ϕ, was a (d+1) x 1 vector of only 0s. The third was A, which was an N x (d+1) matrix where every row consisted of $-y_n$ and $-y_n x_n^T$ going through every element in X. The last parameter was c, which was an N x 1 vector of -1s. After getting all four parameters, they had to be entered as quadprog(H, ϕ, A, c) and set to a variable q. Variable q ended up being a vector of (d+1) x 1 where the first element of the vector was b and the rest of the elements made up w. To find the margin, a new variable had to be set to $1/\|w\|_2$. The decision boundary plus the margin and the decision boundary minus the margin made up the other two lines along the support vectors. Lastly was to find and label the support vectors. Going through each element in X, if $-y_n(w^T x_n + b) = 1$, then that element is considered a support vector and would be labelled accordingly.

Experiments:



Purple Line: Decision boundary
Black Lines: Lines along support vectors
Green Highlighted Coordinates: Support vectors

Support Vector Coordinate 1: (0.760000, 0.750000)
Support Vector Coordinate 2: (0.080000, 0.050000)
Support Vector Coordinate 3: (0.690000, 0.750000)
Margin: 0.025105

Discussion:
As shown in the graph, the lines along the support vectors are resting against the green highlighted coordinates, which are labeled as the support vectors. The margin found for this set of data was 0.025105, which is the largest distance that the closest points on either side of the boundary will be equidistant from it. There were three points that supported these lines and are mentioned above in the experiment section. Like said in the introduction, this will allow for more accurate data since any other data entered can be correctly categorized when the decision boundary is in the most suitable position. This also makes more sense since the decision boundary line should be right in the middle of the closest data points instead of favoring one side.