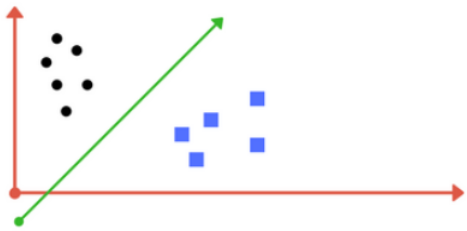
In [8]:
```
#support vector machines basically are used for cassification.It is supervised machine learning algorithm
#reference:
#https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72
#https://medium.com/coinmonks/support-vector-machines-svm-b2b433419d73
```

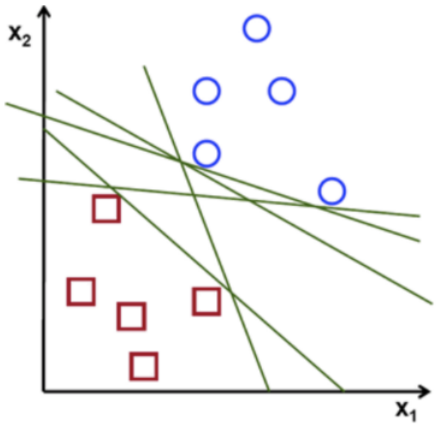
In [9]:
```
#Regression means response is quantitative or numerical value.
#classification means response is categorical/fixed value.
```

In [12]:
```
#To classify the target variable, SVM creates a hyper-plane. (In 2 dimensions, hyperplane is just a line, In 1D its a dot)
```

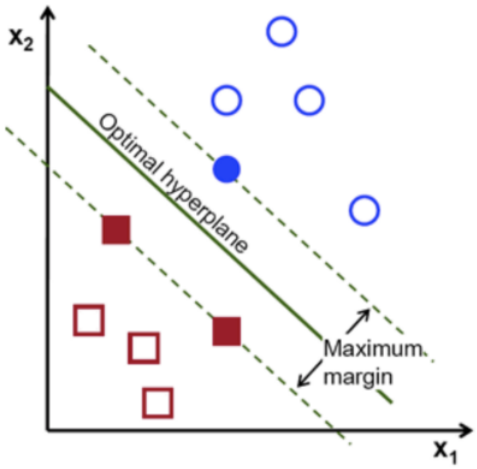


In the graph, there clearly exists two classes. One class is represented with blue circles and another class with red squares. Now we have to find a line that seperates the two classes. As you can see there as many lines that seperates the two classes, so what is the optimal line in this case?

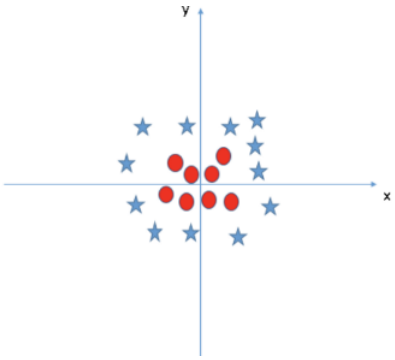
A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal should be to find the line passing as far as possible from all points .



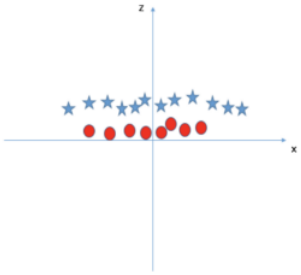
SVM algorithm work by fitting the hyperplane whcih has the largest minimum distance to the training examples. Twice this distance is called margin. Therefore, the optimal separating hyperplane maximizes the margin of the training data.



WHAT IF DATAPOINTS ARE NOT LINEARLY SEPERABLE?



As you can see there does not exist a line that can separate the two classes. SVM can solve this problem. We can introduce a new feature $z=x^2 + y^2$ and plot the z feature.



SVM has a technique called the kernel trick. These are functions which takes low dimensional input space and transform it to a higher dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels. It is mostly useful in non-linear separation problem.

In [ ]: