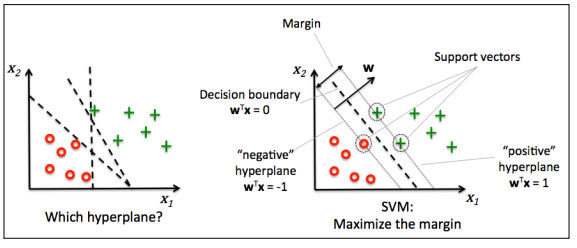
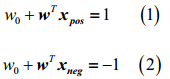
**SVM (support vector machine):**

It can be considered as an extension of the perceptron. Using the perceptron algorithm, we minimized misclassification errors. However, in SVMs, our optimization objective is to maximize the margin. The margin is defined as the distance between the separating hyperplane (decision boundary) and the training samples that are closest to this hyperplane, which are the so-called support vectors.



**Maximum margin intuition:**

The rationale behind having decision boundaries with large margins is that they tend to have a lower generalization error whereas models with small margins are more prone to overftting. To get an intuition for the margin maximization, let's take a closer look at those positive and negative hyperplanes that are parallel to the decision boundary, which can be expressed as follows:



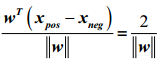
If we subtract those two linear equations (1) and (2) from each other, we get：



We can normalize this by the length of the vector w, which is defined as follows:

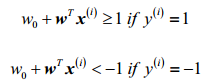


So we arrive at the following equation:



The left side of the preceding equation can then be interpreted as the distance between the positive and negative hyperplane, which is the so-called margin that we want to maximize.

Now the objective function of the SVM becomes the maximization of this margin by maximizing under the constraint that the samples are classified correctly, which can be written as follows:



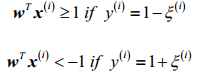
These two equations basically say that all negative samples should fall on one side of the negative hyperplane, whereas all the positive samples should fall behind the positive hyperplane. This can be written more compactly as follows:



**Dealing with the nonlinearly separable case using slack variables**

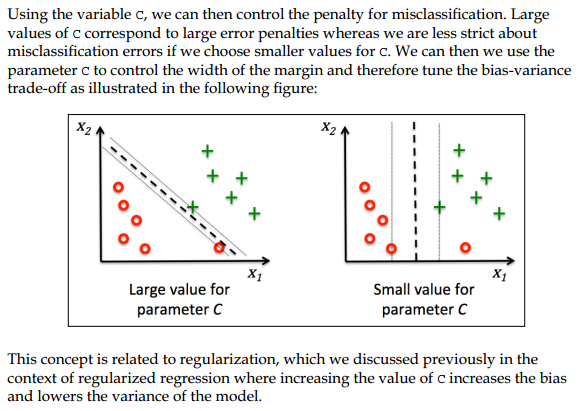
The motivation for introducing the slack variable was that the linear constraints need to be relaxed for nonlinearly separable data to allow convergence of the optimization in the presence of misclassifications under the appropriate cost penalization.

The positive-values slack variable is simply added to the linear constraints:



So the new objective to be minimized becomes:





Reference:

1. Python machine learning