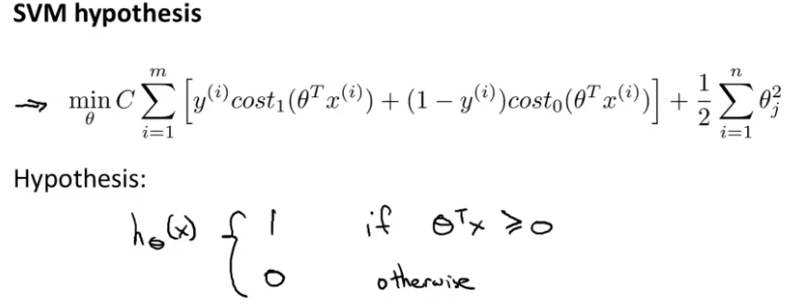
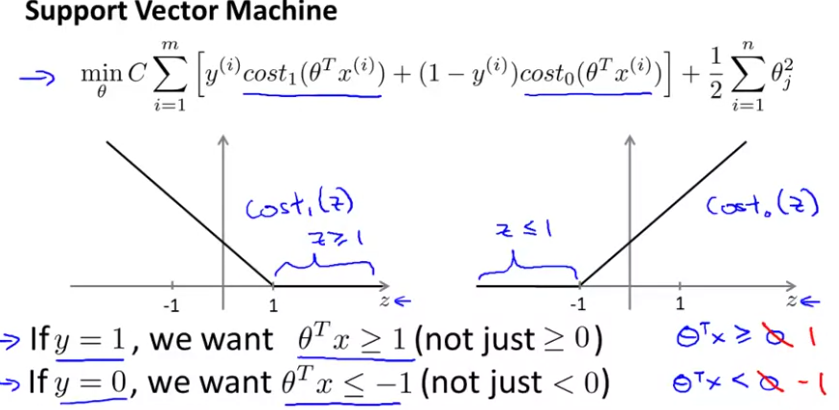
**Optimization objective**

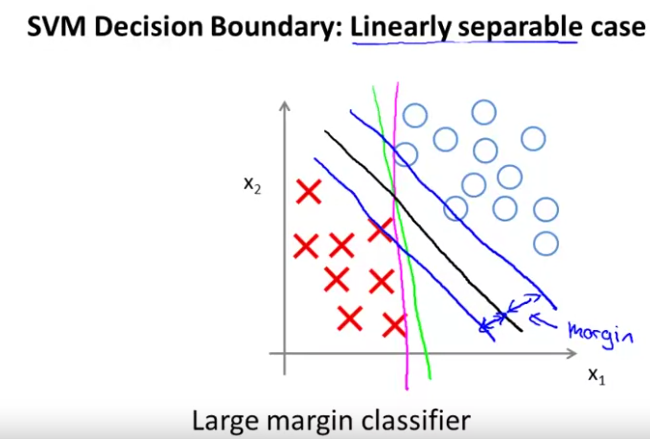
The support vector machine consists in taking the logistic’s regression cost function and approximating it by linear functions. Then, the optimization objective becomes:



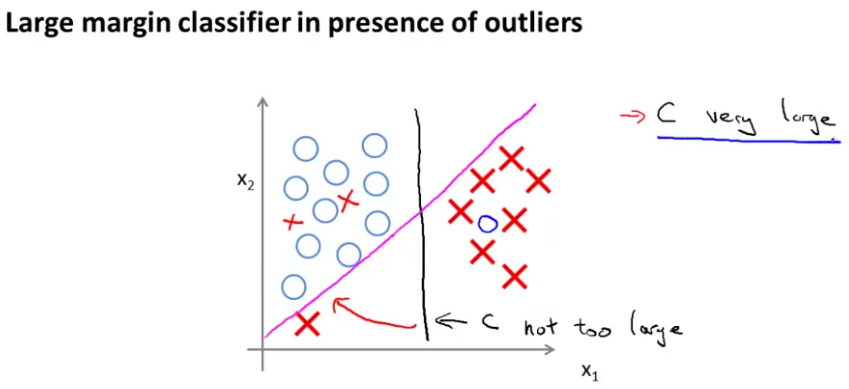


In SWM, we want to include a safety margin factor

By adding this safety margin factor, the SVM becomes a quite more robust algorithm.

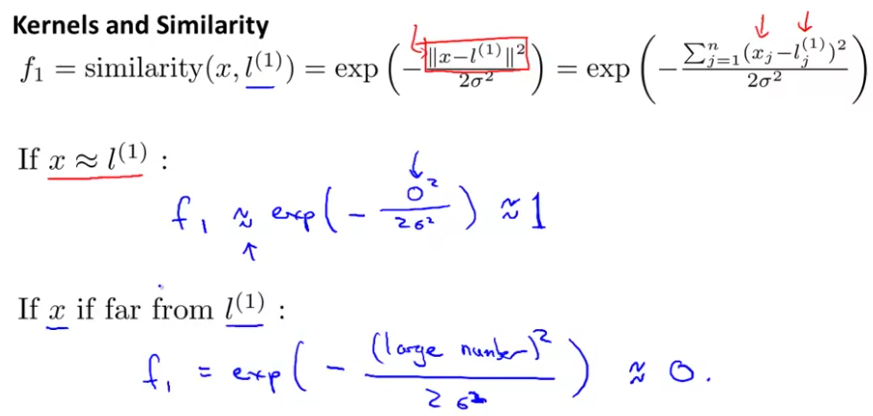


With the parameter C, we can also change the margins by changing the shape of the boundary.

****

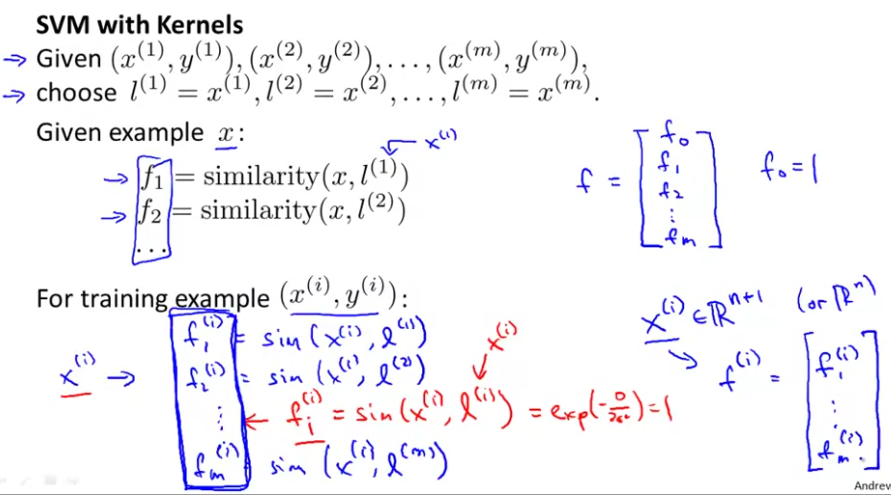
**Kernels**

They allow using SVM to develop non-linear boundaries. For doing so, we define new features from the data and some other points called landmarks using the kernels (this example is for a Gaussian Kernel):

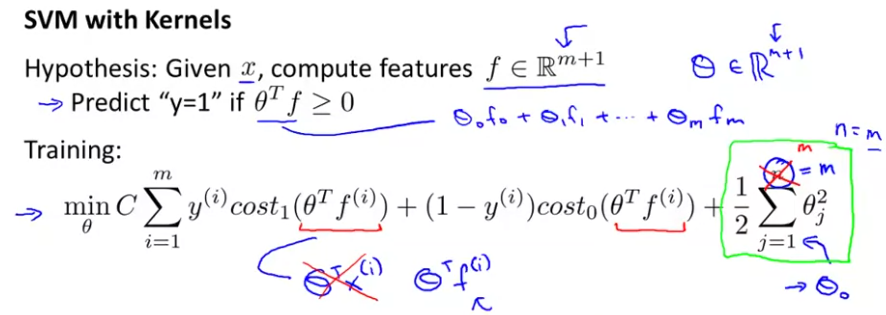


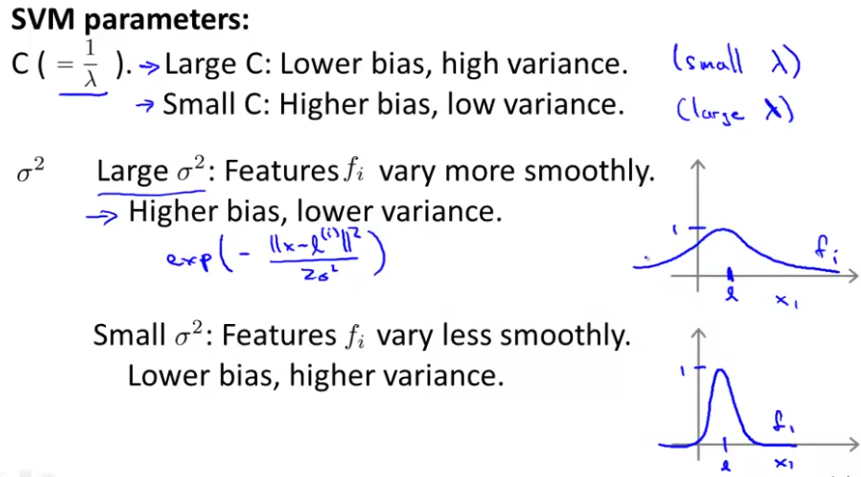
This similarity functions can give us information about how close are x and l, as they are equal to 1 when both points are close, and equal to 0 if they are far away. We can see that if sigma is bigger, f1 gets to 0 when x and l are further.

For choosing the landmarks:

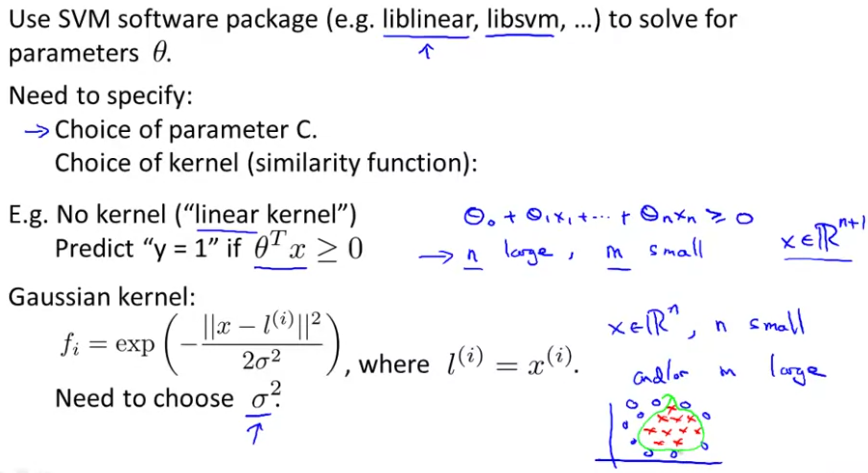


Using kernels with the SVM algorithm :

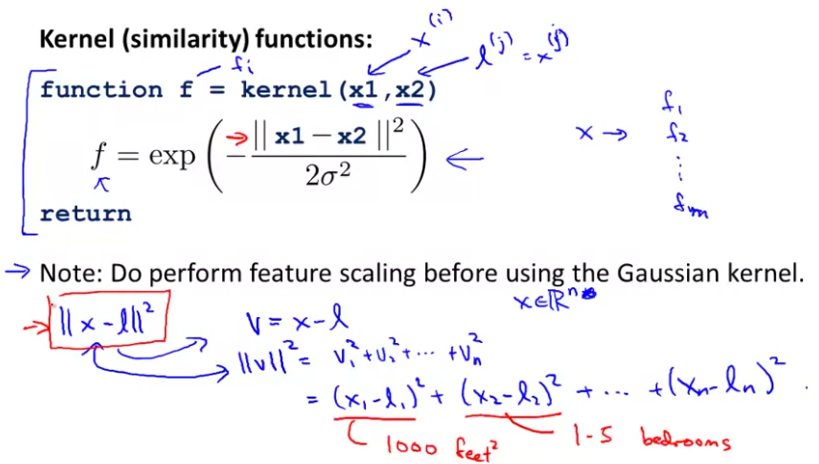
How To choose the SVM parameters:

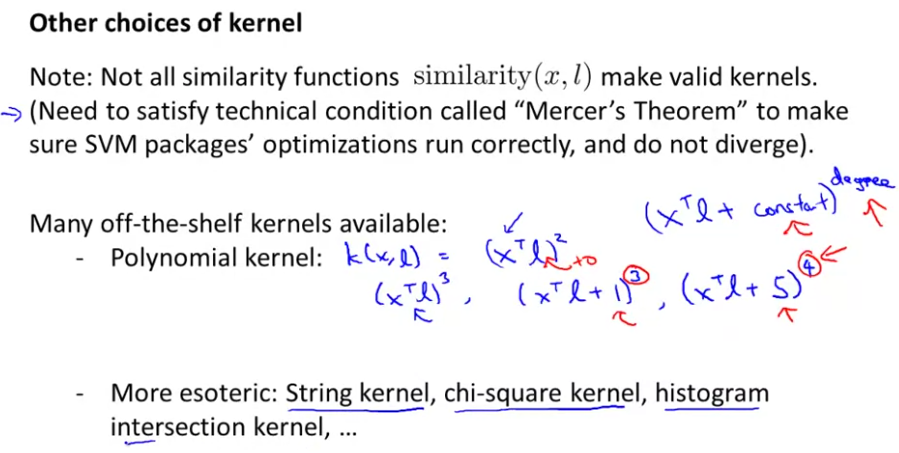


**How to use a SVM**

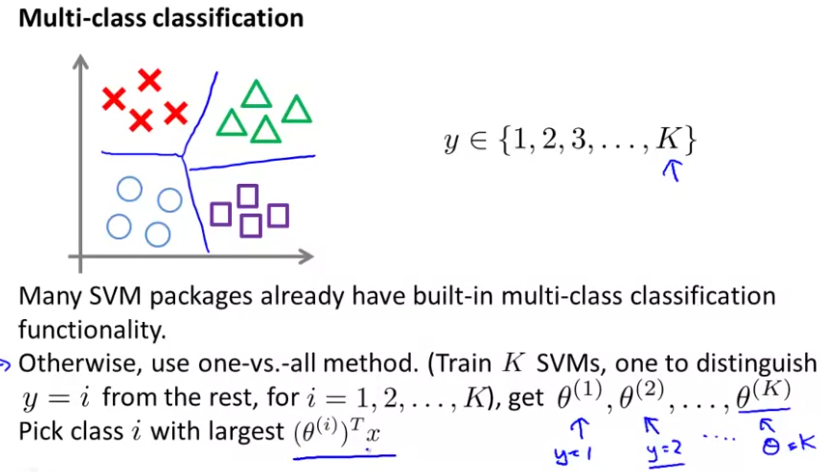


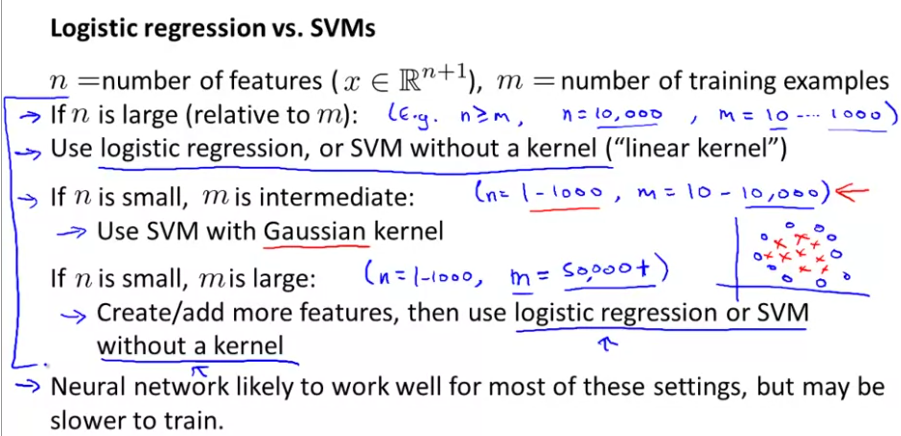
C = 1/λ. Variance vs bias compromiso.





For text classification problems





The optimization problem that the SVM has is a convex optimization problem and so the good SVM optimization software packages will always find the global minimum or something close to it. And so for the SVM you don't need to worry about local optima.