Discussion 5
Machine Learning,Spring 2018

1 KKT Conditions

1.1 Inequality Constraint

Let $\alpha \in \mathbb{R}$ and $a \in \mathbb{R}^n$ with $a \neq 0$. Define the halfspace $H = \{x \in \mathbb{R}^n : a^T x + \alpha \geq 0\}$. Consider the problem of finding the point in H with the smallest Euclidean norm.

(a) Formulate this problem as a constrained optimization problem.

(b) Solve the problem with the help of the KKT conditions. (Hint: you should consider different cases based on if α is negative or nonnegative.)

2 SVM

2.1 Concepts

(a) True or False: The maximum margin hyperplane is only defined by the location of the support vectors, thus other data points can be moved around freely (so long as they remain outside the margin region) without changing the decision boundary. Explain.

(b) True or False: Suppose there is a data set just containing two data points from different classes. If we fit a SVM, then there is a unique solution for the location of the maximum margin hyperplane. Explain.

2.2 Practice

Recall the soft margin SVM primal problem

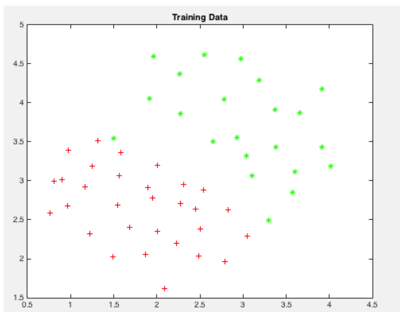
$$\min\left(\frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w} + C\sum_{i=1}^{n} s_i\right), \quad \forall i = 1,\dots,n, \quad s_i \geq 0, \quad (\boldsymbol{w}\cdot\boldsymbol{x}_i + b)y_i \geq (1 - s_i)$$

For the hard-margin case, we have $s_i = 0, \forall i$. Varying the C parameter changes the position of the decision boundary. We can obtain the kernel SVM by considering the dual and replacing $\boldsymbol{x}_i \cdot \boldsymbol{x}_j$ with $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

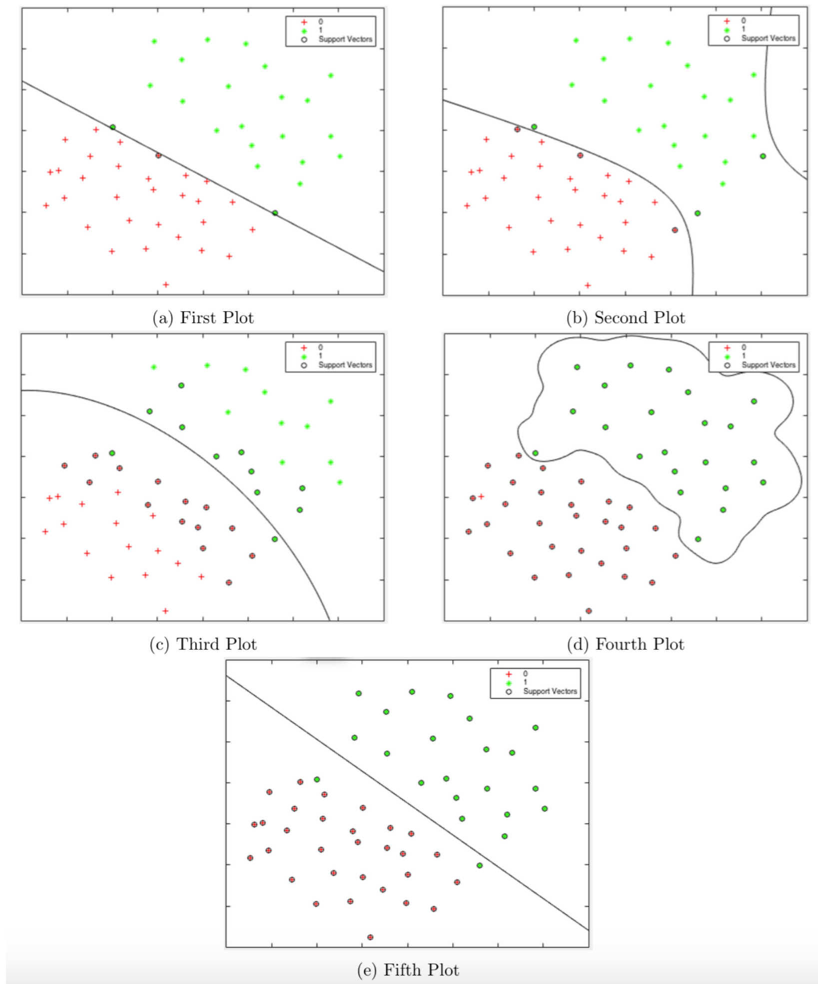
Consider the following dataset of points with two classes, shown in Fig. 1 (a). We decided to play around with the C parameter and kernel functions and observed how the decision boundary changes. We plotted a couple of pictures (shown in Fig. 1 (b)). However, we were careless and didn't label the plots! Answer the following questions:

(a) Select the plot that best corresponds to a soft-margin linear SVM with $C = 0.01$.

(b) Select the plot that best corresponds to a soft-margin linear SVM with $C = 100000$.



(a) Data in a 2D space.



(a) First Plot

(b) Second Plot

(c) Third Plot

(d) Fourth Plot

(e) Fifth Plot

(b) Testing

Figure 1: Data for Decision Trees.

(c) Suppose we were given information that the data we have might be slightly noisy: a couple of points are moved a short distance away from where they actually should be. Which of the above two classifiers do you think does better in this case?

(d) Say you are allowed to add a new datapoint to the dataset. State briefly that in which cases the decision boundary will not change when using the classifier that you selected for the previous part?

(e) Select the plot that best corresponds to a soft-margin quadratic kernel SVM with $C = 0.1$.

(f) Select the plot that best corresponds to a soft-margin quadratic kernel SVM with $C = 100000$.

(g) Select the plot that best corresponds to a soft-margin gaussian kernel ($\sigma = 0.2$) SVM with $C = 10000$.