

CS 189 Spring 2016: Introduction to Machine Learning - Discussion 8

1. Perceptron and SVM Kernel

We are given a set of n training points $\{\mathbf{x}^{(i)}\}$ and associated training labels $\{y^{(i)}\}$. On this data, we train a perceptron with a Gaussian kernel,

$$K(\mathbf{u}, \mathbf{v}) = e^{-\frac{(\mathbf{u}-\mathbf{v})^2}{2\sigma^2}}$$

using the dual perceptron algorithm. This gives us a set of dual weights $\{\alpha_i\}$.

a) What is the interpretation of the dual weights $\{\alpha_i\}$?

b) How do we classify a new test point \mathbf{x} ?

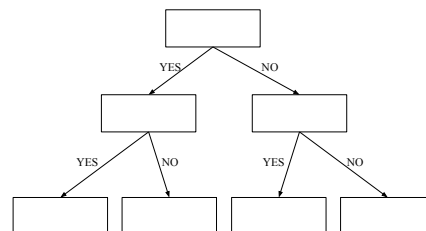
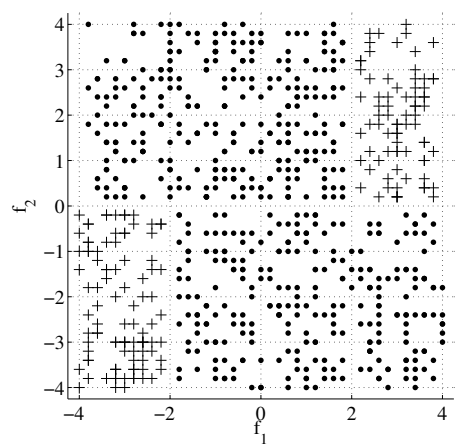
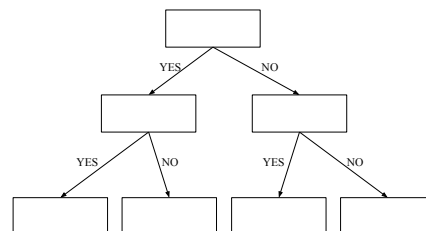
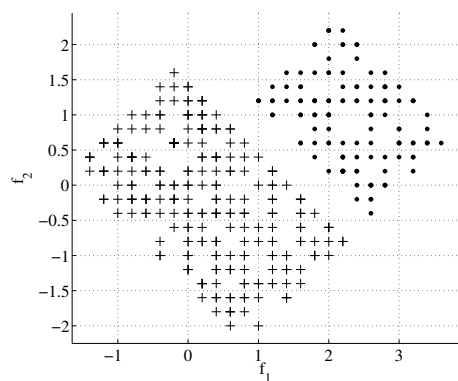
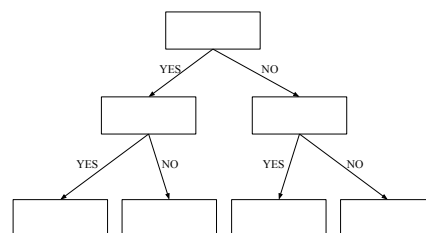
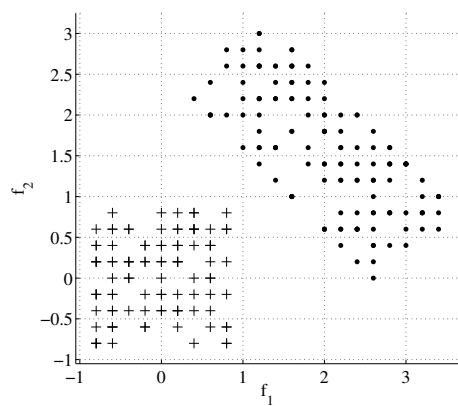
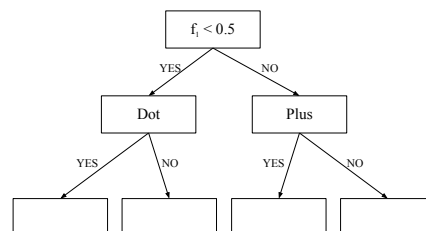
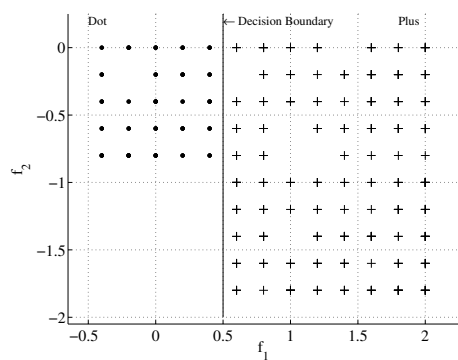
c) Suppose we are training an SVM instead of a perceptron. What changes about the answer to b)?

2. Decision Trees

You are given points from 2 classes, shown as +’s and ·’s. For each of the following sets of points,

1. Draw the decision tree of depth at most 2 that can separate the given data completely, by filling in binary predicates (which only involve thresholding of a *single* variable) in the boxes for the decision trees below. If the data is already separated when you hit a box, simply write the class, and leave the sub-tree hanging from that box empty.
2. Draw the corresponding decision boundaries on the scatter plot, and write the class labels for each of the resulting bins somewhere inside the resulting bins.

If the data cannot be separated completely by a depth 2 decision tree, simply cross out the tree template. We solve the first part as an example.



3. Curse of Dimensionality

We have a training set: $(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(n)}, y^{(n)})$, with $\mathbf{x}^{(i)} \in \mathbb{R}^d$. Our 1-nearest neighbor classifier for a test point \mathbf{x} is:

$$\text{class}(\mathbf{x}) = y^{(i^*)} \quad \text{where } \mathbf{x}^{(i^*)} \text{ is the nearest neighbor of } \mathbf{x}.$$

Assume any data point \mathbf{x} is inside the Euclidean ball of radius 1, i.e. $\|\mathbf{x}\|_2 \leq 1$. To be confident in our prediction, we want the distance between \mathbf{x} and its nearest neighbor to be small, within some positive ϵ :

$$\|\mathbf{x} - \mathbf{x}^{(i^*)}\|_2 \leq \epsilon \quad \text{for all } \|\mathbf{x}\|_2 \leq 1. \quad (1)$$

For this condition to hold, at least how many data points should be in the training set? How does this lower bound depend on the dimension d ?

Hint: Think about the volumes of the hyperspheres, and use the union bound:

$$\text{vol}\left(\bigcup_{j=1}^k S_j\right) \leq \sum_{j=1}^k \text{vol}(S_j), \quad \text{where } S_j \text{ is a hypersphere.}$$