COL 774 - Practice Questions

Tuesday March 22, 2015

Notes:

1. Given an unlabeled set of examples $\{x^{(1)}, \cdots, x^{(m)}\}$ the *one-class SVM algorithm* tries to maximally separate the data from a hyperplane passing through the origin. More precisely, it solves the following (primal) optimization problem:

$$min_w \frac{1}{2} w^T w$$

$$(w^T x^{(i)}) \geq 1, \forall i = 1, \cdots, m$$

Derive the dual form for the above optimization problem. Your dual formulation should not contain any primal variables (w). You should simplify your formulation as much as possible.

2. One way to avoid overfitting in decision trees is to prune the tree using a separate validation set. Typically, a full-blown tree is learnt on the training set first. This is followed by iterative pruning of the learned tree until further pruning does not lead to decrease in error on the validation set. An alternative approach is to keep checking error on the validation set while the tree is being constructed. The tree construction is stopped when the error (on the validation set) does not decrease any further. Which of these approaches do you think would work better in general. Why?

3. Consider a machine learning problem with input feature vectors $x \in R^n$ and satisfying $||x|| = 1$. Given two vectors $x, z \in \mathcal{R}^n$ (and satisfying above properties), consider the function $K(x, z)$ defined as $K(x, z) = ||x + z||^2$. Show that $K(x, z)$ is a valid Kernel.

4. Draw the decision tree of the smallest height to correctly represent the concept in Figure 1. You are allowed to make only two-way splits over an attribute value i.e. any internal node of the tree will have two children. Further, the only decisions allowed are of the form $X < a$, $X > b$, $Y < c$ and $Y > d$, where $a, b, c, d \in \mathcal{R}$. Make sure to indicate on each branch of the tree whether it corresponds to the condition being $true$ or $false$. Also, label each leaf node of the tree appropriately.

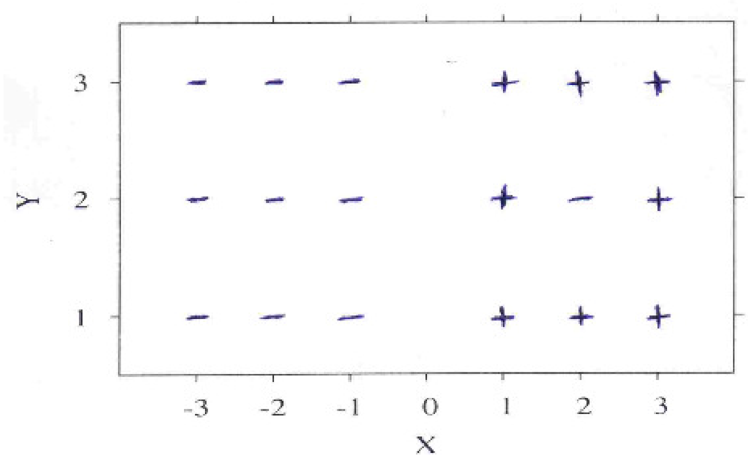

Figure 1: Set of points in two dimensions with corresponding labels.

5. Draw a neural network to represent the Boolean function $f(x_1, x_2, x_3) = (x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee \neg x_3)$ defined over 3 input variables. Note that $x_1, x_2, x_3 \in \{0, 1\}$. Here, $\wedge$ denotes the *and* operator, $\vee$ denotes the *or* operator and $\neg$ denotes the *negation*, as in the standard Boolean algebra. Your network should have at most one hidden layer and use at most 3 network units. Use the threshold function $g(x) = \mathbb{1}\{x \geq 0\}$ ($\mathbb{1}$ represents the indicator function) to process the output of each of the units. Clearly specify the interconnections and the associated weights. Also argue briefly why your construction is correct.