Ariel University

Machine Learning

Homework 2

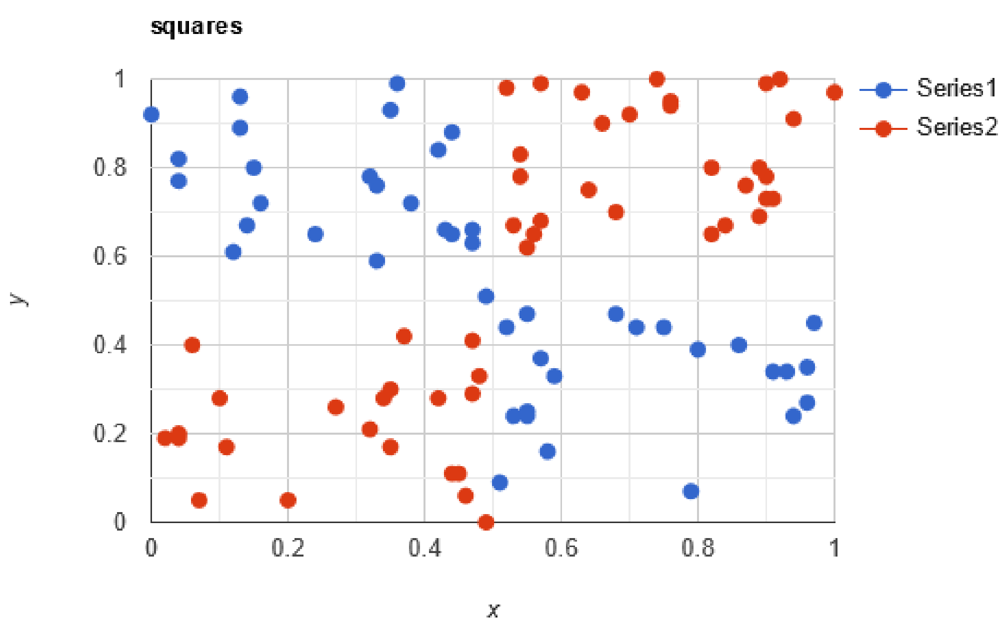
Problem 1.

a) What is the VC-dimension of the infinite set of (uni-directional) balls on three dimensional points? On d-dimensional points? Prove your answer.
b) How many different labels can 4-dimensional (uni-directional) balls give to 100 points? Give an upper bound.

Problem 2.

a) In the class slides, we gave a version of Perceptron that finds a separating hyperplane with margin $\frac{\gamma}{2}$. Give a bound on the number of mistakes made by this algorithm.
b) What happens if Perceptron is run on a set that can't be separated by a hyperplane?

Problem 3. The "squares" data set contains 100 2-dimensional points, where the last column in the file is the labels:



Each pair of points define a line that passes through them. The set of all such lines is our set of rules. Implement Adaboost using these rules.

One run of Adaboost is as follows: Split the data randomly into ½ test (T) and ½ train (S). Use the points of S (not T) to define the hypothesis set of lines. Run Adaboost on S to identify the

8 most important lines h_i and their respective weights α_i. For each k=1,…,8, compute the empirical error of the function H_k on S, and the true error of H_k on T:

$$H_k(x) = sign(\sum_{i=1}^{k} \alpha_i h_i(x))$$

$$\bar{e}(H_k) = \frac{1}{n} \sum_{x_i \in S} [y_i \neq H_k(x)]$$

$$e(H_k) = \frac{1}{n} \sum_{x_i \in T} [y_i \neq H_k(x)]$$

Execute 50 runs of Adaboost, and report $\bar{e}(H_k)$ and $e(H_k)$ for each k, averaged over the 50 runs. Hand in printouts of the values of $\bar{e}(H_k)$ and $e(H_k)$ (total: 16 values). Answer the following:

1. Analyze the behavior of Adaboost on S and T. Do you see any exceptional behavior? Explain.

2. Do you see overfitting? Explain.