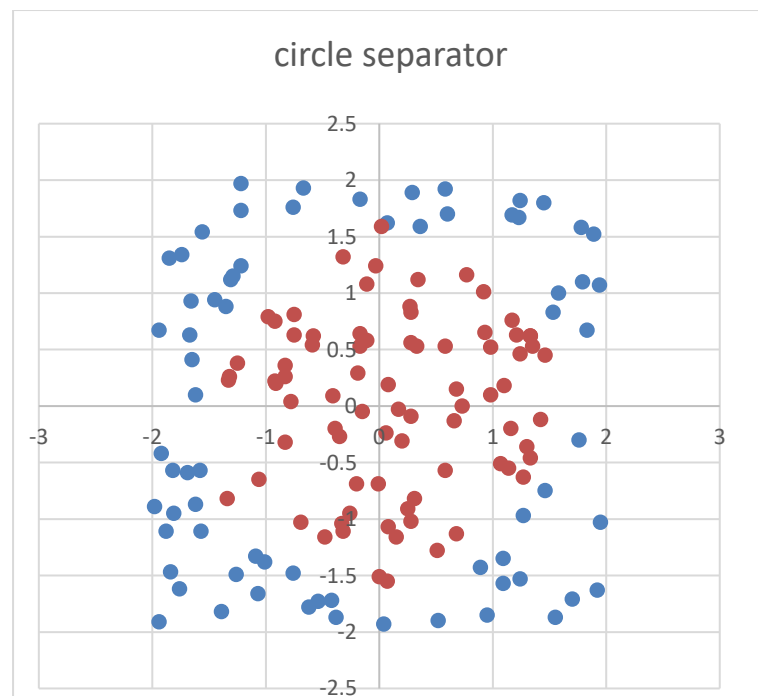


Problem 1.

- a) What is the VC-dimension of the infinite set of uni-directional balls on three dimensional points? Prove your result.
- b) Hypothesis class **C** contains both the infinite set of uni-directional balls and the infinite set of hyperplanes on three dimensional points. What is its VC-dimension? Prove your result.
- c) Hypothesis class **D** contains the infinite set of half-balls on three dimension points. Give an upper bound on its VC-dimension.
- d) How many different labels can **D** give to 100 points? Give an upper bound.



Problem 2. The “circle separator” data set contains 100 2-dimensional points, where the last column in the file is the label:



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 $\frac{1}{2}$ test (T) and $\frac{1}{2}$ 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 $k=1,\dots,8$, compute the empirical error of the function H_k on S, and the true error of H_k on T:

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

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

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

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 $e(H_k)$ for each k (total: 16 values).

Answer the following:

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

B. Do you see overfitting? Explain.

Now repeat the above experiment with circles instead of lines. Two points can define a circle, with one point being its center and the other giving the radius.

C. Answer questions A&B for circles.

D. How do circles compare with lines? Why is this?