MACHINE INTELLIGENCE AND EXPERT SYSTEMS
EC60091
AUTUMN SEMESTER - 2018

ASSIGNMENT-4

(Metrics, VC Dimension, Feature Selection and Hypothesis Evaluation)

1) Prove that VC(H) ≤ $\log_2|H|$ where H is a hypothesis space.
   |H|=Total number of hypothesis. VC(H) implies VC-Dimension of H.
   A hypothesis on a set of n points, defines which of two classes(positive/negative output class) can each point belong to. For example, in a rectangle hypothesis all points inside it can be labelled positive and the others negative. A hypothesis space is a family of all possible hypotheses.

2) Consider the instance space X = {1, 2, ..., 999}. Let H be a hypothesis class consisting of 10 hypothesis, h0 through h9. A number n ∈ X is an element of hi if the normal decimal representation of n contains the digit i. So, for example, the number ”778” is an element of h7 and h8. What is the VC dimension of H? Justify your answer.
   (Hint: To prove the VC dimension of H is k, you should justify the following claims.
   1. Use the property proved in the previous question to get an upper bound
   2. Generate a possible lower bound)
3) ThirdEye Technologies have innovated a low-cost solution for cancer detection catering to India’s rural demand. You have been sent to analyze their product in a cancer-specialty hospital having 1000 patients. Based on hospital’s data, 200 patients are suffering from cancer. The device detects 160 of them with cancer while the other cancerous patients are declared free of disease. 10 people without cancer are detected with cancer while the rest are not.
   i) Your objective is to create a confusion matrix to study the effectiveness of the device.
   ii) Based on the above confusion matrix calculate the following metrics for the device: Accuracy, Precision, Recall, Sensitivity, Specificity, F1-Score.