

PATTERN RECOGNITION AND CLASSIFICATION

WEEK 1 ASSIGNMENT ANSWERS

ARYA RAJIV CHALOLI

PES1201700253

1. What do you understand by Pattern recognition? Why do we need it?

Answer:

Pattern recognition is the act of taking in raw data and taking an action based on the “category” of the pattern.

From automated speech recognition, fingerprint identification, optical character recognition, DNA sequence identification etc, a reliable, accurate pattern recognition by machine would be immensely useful.

2. What do you mean by decision cost? Give an example where making a choice between two options would have equal cost and an example where the cost would be different.

Answer:

Decision cost is primarily the “consequence” of a decision and the action taken that follows (The consequence could be the loss incurred if the classifier erroneously classifies an object).

A case when the consequences of our actions are equally costly: if the classifier is classifying pens according to the colour, the mis-happening of classifying a blue pen to the class of black pens is just as costly as classifying a black pen to class of blue pens.

On the contrary all cases are not as symmetrical as the case mentioned above. For example: While classifying salmon and seabass', if a salmon is wrongly classified into the set of seabass', the customer would not object to the fact as much as the condition in which a seabass is classified into a set of salmon. Hence in this case the cost of an error is unsymmetrical.

3. Why do we need generalization in our classifiers?

Answer:

Generalization is necessary because the central aim of designing a classifier is to suggest actions when presented with novel patterns.

Generalization is the process of generalizing or simplifying the base model based on which the decision is taken. An overtly complex model/decision boundary would not provide us the true model as the it would seem to “tuned” for the particular training samples and would prove to be not very effective in the case of a novel patterns. Hence generalization is essential.

4. What factors should we keep in mind when selecting features for our classifier?

Answer:

Firstly, the feature we choose must be the most evident feature that distinguishes between the two or more categories that we intend to classify.

But in certain cases, even if we know the costs associated with our decisions and choose the optimal decision boundary, we may be dissatisfied with the resulting performance. In such cases we move on to the next most evident feature. But it should always be retained in mind that, the inclusion of too many features would create an overtly complex model and fail the requirement of sorting novel patterns. Hence, we have to optimally choose our features, such we don't compromise heavily on the accuracy, nor to create an overtly specialised model, tuned to our training sets.

5. Calculate a patient's probability of having liver disease if they are an alcoholic. A is the event “Patient has liver disease.” Past data tells you that 10% of patients entering your clinic have liver disease. $P(A) = 0.10$. B is the event that that “Patient is an alcoholic.” Five percent of the clinic's patients are alcoholics. $P(B) = 0.05$. You might also know that among those patients diagnosed with liver disease, 7% are alcoholics.

Answer:

Ans: 0.14

$$P(A|B) = (P(B|A) * P(A)) / (P(B))$$

$$P(A|B) = (0.07 * 0.1) / (0.05) = 0.14$$