Weekly report of lessons

Name: Suryam Arnav Kalra
Roll No: 19CS30050
The week: 13th September 2021 to 17th September 2021

The topics covered:

- Optimal Classifier
- Gibbs Algorithm
- Discriminant Function
- Challenges in Computing
- Naïve Bayes Classifier and Likelihood estimation in Naïve Bayes
- Avoiding zero-probability and Pros and Cons of Naïve Bayes
- Bayesian Network and Formation of a Graphical model
- Conditional Independence
- Computation on Bayesian Network and Inference through Bayesian Network
- Losses and Risks in Bayesian Decision Making and Optimum Classification Rule
- Mining association Rules
- Apriori Algorithm
- Association and Causality

Summary topic wise:

- Optimal Classifier: Learning the target function of a classifier as $c(x) = argmax\ \sum P(v|h)P(h|D)$
- Gibbs Algorithm: Instead of enumerating exhaustively choose a hypothesis h randomly for an instance x with posterior distribution $P(h|D)$ and apply h on x.
- Discriminant function: Bayesian Classifiers can be expresses in the framework of a set of discriminant functions $g_i(x)$ and assign $C_i\ if\ g_i(x) > \ g_k(x)\ for\ all\ k\ (exc. i)$
- Challenges in computing: We need to have prior knowledge of probabilities of classes along with probability distributions in multidimensional feature spaces.
- Naïve Bayes Classifier: It makes the naïve assumption that the attributes are conditionally independent.
$$P(X|C_i) = \prod P(\,x_k|C_i)$$
- Likelihood estimation in Naïve Bayes: We can use fraction of times the value occurs for a discrete variable and can use Gaussian distribution for continuous variable.
- Avoiding Zero-Probability: To avoid zero-probability we use **Laplacian correction** where we add 1 to each case to avoid making the likelihood zero.
- Pros and Cons of Naïve Bayes: It is easy to implement and gets good results in most of the cases. But, due to a naïve assumption it leads to a loss of accuracy since in real life, dependencies exist among variables which cannot be modeled by Naïve Bayes.
- Bayesian Network: It is a more general framework for modeling conditional dependencies:
    - It is composed of nodes and arcs between the nodes.
    - It makes a directed acyclic graph (DAG)
    - Topology called structure and P(X), P(Y|X) are parameters.

- Formation of a graphical model: Form a graph by adding nodes and arcs between two nodes, if they are not independent. X and Y are independent if P(Y|X) = P(Y) and P(X|Y) = P(X).
- Conditional Independence: Conditional independence between X and Y given the occurrence of Z can be written as P(X,Y|Z) = P(X|Z)P(Y|Z)
- Computation on Bayesian Network: Given the value of any set of variables as an evidence infer the probabilities of any other set of variables.
- Inference through Bayesian Network: Given any subset of X_i , calculate the probability distribution of some other subset of X_i by marginalizing over the joint.
- Losses and Risks in Bayesian Decision Making: Let a_i be the ith action of assigning x to class C_i and let l_{ik}be the loss due to a_i if x belongs to C_i then the expected risk for taking this action is $R(a_i|x) = \sum l_{ik} P(C_k|x)$ and we choose a_i which minimizes R(.)
- Optimum Classification Rule: Choose a_i if $P(C_i|x)$ is maximum among i = 1,2,…,K and else reject (No class assignment).
- Mining Association Rules: An association rule is an implication X -> Y where X is the antecedent and Y is the consequent.
    - Support(X, Y): P(X, Y) = # of customers who bought X and Y / # of total customers.
    - Confidence(X -> Y): P(Y|X) = P(X, Y)/P(X) = # of customers who bought X and Y/ # of customers who bought X.
    - Lift(X, Y) = P(X, Y)/P(X)P(Y) = P(Y|X) / P(Y)
- Apriori Algorithm: This algorithm is used to get association rules with high support and confidence from a database.
    - Step 1: Find frequent item sets, that is, those which have enough support.
    - Step 2: Convert them to rules with enough confidence, and in every pass reduce antecedent part and increase the consequent part.
- Association and Causality: X -> Y indicates association and not causality.

Concepts challenging to comprehend:

Discriminant function along with Bayesian Network are a little bit challenging to comprehend.

Interesting and exciting concepts:

Naïve Bayes Classifier along with the representation of the problem as a graph are quite interesting and exciting to learn.

Concepts not understood:

After going through the book and the video lectures the concepts are clearly understood.

Any novel idea of yours out of the lessons:

Instead of making all the attributes independent in Naïve Bayes, we can consider making a set of attributes independent and making the attributes inside these sets dependent on each other to better model real life scenarios. As an implementation point of view, the product of very large number of probabilities will be too small and not handled correctly, to avoid this we can take the logarithm of the probabilities and add them since log(x.y) = log(x) + log(y).