Weekly report of lessons

Name: Ravi Pratap Singh
Roll No: 20CS60R60
The week: 14 oct ,15 oct, 16 oct

The topics covered : Losses and risks, Mining association rules ,Apriori Algorithm,Dimensionality reduction , Feature selection , Principal Component Analysis (PCA) ,Principal Components.

Summary topic wise :

- Loss function $L(\theta,\ \delta(x))$ is a function of unknown parameter $\theta \in \Theta$. $\delta(x)$ is a decision based on the data The loss function determines the penalty for deciding how well $\delta(x)$ estimates θ .$L(\theta,\ \delta(x))$ = 0 if $\delta(x) = \theta$ and = 1 otherwise . Risk measures the long-term average loss resulting from using δ. Expected Risk for taking action δ is $R(\theta/\delta(x)) = \sum_{k=1}^{k=n} L(\theta, \delta) * P(x/\theta)$. Minimizing the risks implies maximizing posterior .
- Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets . An implication is $X \rightarrow Y$, Where X is antecedent and Y is consequent .
  Support: Support indicates how frequently the if/then relationship appears in the database ⇒ $P(X, Y)$.
  Confidence: Confidence tells about the number of times these relationships have been found to be true ⇒ $P(Y/X) = P(X, Y) / P(X)$. Lift = $P(X, Y) / [\ P(X) * P(Y)\ ] = P(Y/X) * P(Y)$
- Apriori algorithm uses prior knowledge of frequent itemset properties .This algorithm uses two steps “join” and “prune” to reduce the search space. It is an iterative approach to discover the most frequent itemset .Its Disadvantage is that it requires high computation if the itemsets are very large and the minimum support is kept very low.
- Dimensionality reduction refers to techniques that reduce the number of input variables in a dataset. It is required because Large numbers of input features can cause poor performance for machine learning algorithms .It include feature selection, linear algebra methods and projection methods.
- Feature selection is the process of selecting a subset of relevant features for use in model construction .
  In Sequential forward selection we start with empty set $F = \phi$, then select the next best feature which provide least $E(F \cup x_i)$, update $x_i\ if\ E(F \cup x_i) < E(F)$.
  In Sequential backward selection First, the criterion function is computed for all n features .Then, select x_i ,which provide least $E(F - x_i)$.Next,update $x_i\ if\ E(F \cup x_i) < E(F)$. This procedure continues until no more removal is possible.
- PCA uses simple matrix operations to calculate a projection of the original data into the same number or fewer dimensions. Advantage : It is reduction of noise since the maximum variation basis is chosen and so the small variations in the background are ignored automatically. Disadvantages :simplest invariance could not be captured by the PCA unless the training data explicitly provides this information.
- The first principal component is the direction w_1 in space along which projections have the largest variance. The second principal component is the direction which maximizes variance among all directions orthogonal to the first. The kth component is the variance-maximizing direction orthogonal to the previous k − 1 components.
  Random Variable for first component is , $z_1 = {w_1}^T x$. To maximize variance keeping w_1 as unit vector as $w_1 = argmax_w\ \{w^T \sum w - l * (w^T w - 1)\}$ ⇒ taking derivation , $\sum w_1 = l\, w_1$. w_1 is eigenvector of $\sum$ corresponding to maximum eigenvalue. Similarly w_2 could be calculated with the help of fact that w_2 is orthogonal to w_1

Concepts challenging to comprehend : Mathematics behind principal components
Any novel idea of yours out of the lessons : Apriori algorithm could be combined with the cause of one disease with another and predict how likely the person is vulnerable to if he has had some disease or in contact with someone having it.
Was the last doubt clearing session useful? : Yes