| Question? | Sources | | | | | | | |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jadhav, S. D., & Channe, H. P. (2016). Compar. | Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning | Lotte, F., Bougrain, L., Cichocki,i | Mr.Y.Madhusekhar, Dr G. Nanda Kishore kumar Dr K. Umap | Muhamedyev, R., Yakunin, K., Iskakov, S., Sainova | Naik, A., & Samant, L. (2016). Correlatio. | Reddy, R. V. K., & Babu, D. U. R. (2018). A Review on Classification | Zhang, C., Liu, C., Zhang, X., & Almpanid | Dayton, C. M. (1992). Logistic regression analysis. *Stat*, 474-574. | Sinha, P., & Sinha, P. (2015). Comparative study of chronic kidney disease prediction using KNN and SVM. *International Journal of Engineering Research and Technology*, *4*(12), 608-12. | Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest *Engineering Research and Applications*, *3*(5), 605-610. |  |
| What is classification? | -Classification  is a data mining technique used  to predict group membership for data instances within a given dataset.  -aims  at learning the relationship between a set  of feature variables and a  target variable  of interest  -Classification  is considered  as an  example  of supervised learning  as training data associated with class labels  is given  as input  -classification  is the process  of organizing data into  categories/groups  in such a way that data objects  of same  group are more similar and data objects from different  groups are very dissimilar |  |  | -Classification is a process of setting up the boundary conditions to predict the target class  and provides a classifier so as to determine the possible outcome based on the independent  variables. |  |  |  |  |  |  | -Classification problems aim to identify the characteristics that indicate the group to which each case belongs |  |
| What are the application of classification? | -Customer Target Marketing, Medical Disease  Diagnosis, Social Network Analysis, Credit Card Rating, Artificial Intelligence, and Document Categorization |  |  |  |  |  |  |  |  |  |  |  |
| What is a classification algorithm? | -Classification algorithm assigns  each instance  to a particular class such that classification  error will  be least  -It is used  to extract models that accurately  define important data classes within the given dataset  -It can predi  ct categorical class labels  and classifies data based  on model built  by using training set  and associated class labels and then can  be used for  classifying newly available test data  -Classification process  is divided into two main steps.  Thefirst  is the training step where the classification model  is  built. The second  is the classificati  on itself,  in which the  trained model  is applied  to assign  unknown data object  to  one out  of a given set  of class label |  |  |  |  |  |  |  |  |  |  |  |
| What is the need for classification algorithms? | It is  used for classifying data into different classes  by considering some constrain | -The  resulting classifier is then used to assign class labels to the testing instances where the values  of the predictor features are known, but the value of the class label is unknown. |  |  |  |  |  |  |  |  |  |  |
| How does classification differ from other algorithms? |  |  |  |  |  |  |  |  |  |  |  |  |
| What are the advantages of classification algorithms? |  |  |  | -are more accurate and faster to predict the results  of the huge dataset |  |  |  |  |  |  |  |  |
| What are the disadvantages of classification algorithms? | -a dataset collected by the  “brute-force” method is not directly suitable for induction. It contains in most cases noise  and missing feature values, and therefore requires significant pre-processing |  |  |  |  |  |  |  |  |  |  |  |
| What is Logistic Regression(LR)? |  |  |  |  |  |  |  |  |  |  |  |  |
| Advantages of LR |  |  |  |  |  |  |  |  |  |  |  |  |
| What is Naïve Bayes(BA)? | - Naive Bayes Classifier  is the simple Statistical Bayesian  Classifier  - It is called Naive  as it assumes that all  variables contribute towards classification and are mutually  correlated  - It is also called Idiot’s Bayes, Simple  Bayes, and Independence Bayes  - A Naive  Bayes classifier considers that the presence (or absence)  of a  particular feature(attribute)  of a class  is unrelated  to the  presence (or absence)  of any other feature when the class  variable  is given  - it is used when the dimensionality  of the inputs  is high  The Naive Bayes classifier [2] works  as follows:  1.  Let D  be the training dataset associated with class  labels.Each tuple  is represented  by n-dimensional element  vector, X=(x1, x2, x3,.....,xn).  2.  Consider that there are m classes C1, C2, C3...., Cm.  Suppose that  we want  to classify  an unknown tuple  X,  then the classifier will predict that X belongs  to the class  with higher posterior probability, conditioned  on X. i.e.,  the Naive Bayesian classifier assigns  an unknown tuple X  to the class  Ci if and only  if P(Ci|X) > P(Cj|X) For 1 ≤ j ≤  m, and i≠j, above posterior probabilities are computed  using Bayes Theorem |  |  | -A Naive  Bayes classifier is a simple probabilistic classifier  with strong (naive) independence assumptions  which  based on  applying Bayes' theorem  . In simple terms,  a Naive Bayes  classifier assumes that the presence (or absence) of a particular  feature of a class is unrelated to the presence (or  absence) of  any other feature |  |  |  |  |  |  |  |  |
| Advantages of BA | •  It requires short computational time for training.  •  It improves the classification performance  by removing  the irrelevant features  .  •  It has good performance. |  |  | -advantage of the Naive Bayes classifier is that it only  requires a small amount of t  raining data to estimate the means  and variances of the variables necessary for classification.  Because independent variables are assumed, only the  variances of the variables for each  label  need to be determined  and not the entire covariance matrix |  |  |  |  |  |  |  |  |
| disadvantages of BA | •  The Naive Bayes classifier requires a very large number  of records  to obtain good results.  •  Less accurate  as compared  to other classifiers  on  some  datasets. |  |  |  |  |  |  |  |  |  |  |  |
| What is Stochastic Gradient Descent(SGD)? |  |  |  |  |  |  |  |  |  |  |  |  |
| Advantages of SGD |  |  |  |  |  |  |  |  |  |  |  |  |
| What is K-Nearest Neighbor(KNN)? | -The K-Nearest Neighbor Algorithm  is the simplest  of all  machine learning algorithms  -It is based  on the principle that  the samples that are simila  r, generally lies  in close vicinity  -K-Nearest Neighbor  is instance based learning method  -Lazy-learning algorithms require less  computation time during the training phase than eager-  learning algorithms (such  as decision trees, neural networks  and bayes networks) but more computation time during the  classification process  -Nearest-neighbor classifiers are based  on learning  by  resemblance, i.e.  by comparing a given test sample with the  available training samples which are similar  to it  -If the value  of k is too small, then K-  NNclassifier may  be  vulnerable  to over fitting because  of noise present  in the  training dataset.  On the other hand,  if k is too large, the  nearest-neighbor classifier m  ay misclassify the test sample because its list  of nearest neighbors may contain some data  points that are located far away from its neighborhood.  Initialize value  of K.  2.  Calculate  distance between input sampleand training  samples.  3.  Sort the distances.  4.  Take top K- nearest neighbors.  5.  Apply simple majority.  6.  Predict class label with more neighbors for input sample. |  |  | -The algorithm is based on counting the number of objects  in sphere (hypersphere) of each class centered in a  recognizable object  -the weights of the objects in the sphere can be  described as an inverse proportionally dependence from the  distance to t  he recognizable object. Thus, the closer the object  is, the more important it is for the recognizable object |  |  |  |  |  | -K-Nearest Neighbor algorithm (K-NN) is a non-parametric method used for classification and regression  -In KNN Classification, the output is a class membership  -Classification is done by a majority vote of neighbours    The KNN algorithm goes like this: -initialization, define k -calculate distance between the test sample and all the training samples -sort the distance -take k neighbor -gather the category of nerest neighbor -apply simple majority of category. | -the KNN algorithm is a method for classifying objects based on closest training examples in the feature space  -KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification  -The performance of a KNN classifier is primarily determined by the choice of K as well as the distance metric applied  -If K is very small, the local estimate tends to be very poor owing to the data sparseness and the noisy, ambiguous or mislabeled points  -a large value of K easily makes the estimate over smoothing and the classification performance degrades with the introduction of the outliers from other classes |  |
| Advantages of KNN | •  Easy to understand  and implement.  •  Training  is very fast.  •  It is robust  to noisy training data.  •  It performs well  on applications  in which a sample  can  have many class labels |  |  |  |  |  |  |  |  |  | -The KNN is the fundamental and simplest classification technique when there is little or no prior knowledge about the distribution of the data  -KNN has several main advantages: simplicity, effectiveness, intuitiveness and competitive classification performance in many domains. It is Robust to noisy training data and is effective if the training data is large.  - |  |
| disadvantages of KNN | •  Lazy learners incur expensive computational costs when  the number  of potential neighbors whichto compare a  given unlabeled sample  is large [5].  •  It is sensitive  to the local structure  of the data [1].  •  Memory limitation.  •  As it is supervised lazy learner,  it runs slowly |  |  |  |  |  |  |  |  | -The limitation of the K-NN algorithm is it’s sensitive to the local configuration of the data. | -Distance based learning is not clear which type of distance to use and which attribute to use to produce the best results. Computation cost is quite high because we need  - to compute distance of each query instance to all training samples. |  |
| What is Decision Tree(DT)? | -Decision tree learning uses a decision tree  as a predictive  model which maps observations about  an item  to  conclusions about the item’s target value  - a data mining induction techniques that  recursively partitions a dataset  of records using depth-first  greedy approach  or breadth-first approach until a  ll the data  items belong  to a particular class.  Decision tree classification technique  is performed  in two  phases: tree building and tree pruning [14]. Tree building  is  performed  in top-down approach. During this phase, the tree  is recursively partitioned till all the data items belong  to the  same class label.  It is very computationally intensive  as the  training dataset  is traversed repeatedly. Tree pruning  is done  in a bottom-  up manner.  It is used  to improve the prediction  and classification accuracy  of the algorithm  by minimizing  over-fitting problem  of tree. Over-fitting problem  in decision  tree results  in misclassification error. | -Decision trees are  trees that classify instances by sorting them based on feature values  - Each node in a decision  tree represents a feature in an instance to be classified, and each branch represents a value  that the node can assume. Instances are classified starting at the root node and sorted based  on their feature value.  - The feature that best divides the training data would be the root node of the  Tree  -A decision tree, or any learned hypothesis  h  , is said to overfit training data if another  hypothesis  h  exists that has a larger error than  h  when tested on the training data, but a  smallererrorthan  h  when tested on the entire dataset  -The most straightforward way of tackling overfitting is to pre-prune the  decision tree by not allowing it to grow to its full size |  |  |  |  |  |  |  |  |  |  |
| Advantages of DT | •  Decision Trees are very simple and fast.  •  It produces the accurate result [1].  •  Representation  is easy  to understand i.e. comprehensible.  •  It supports incremental learning [5].  •  It takes the less memory.  •  It can also deal with noisy data.  •  It uses different measures such  as Entropy, Gini index,  Information gain etc.to find best split attribute | -Comprehensibility  - |  |  |  |  |  |  |  |  |  |  |
| disadvantages of DT | •  It has long training time.  •  Decision trees can have significantly more complex  representation for some concepts due  to replication  problem [5].  •  It has a problem  of over fitting |  |  |  |  |  |  |  |  |  |  |  |
| What is Random Forest(RF)? |  |  |  |  |  |  |  |  |  |  |  |  |
| Advantages of RF |  |  |  |  |  |  |  |  |  |  |  |  |
| What is Support Vector Machine(SVM)? |  | Support Vector Machines (SVMs) are the newest supervised machine learning technique |  |  |  |  |  |  |  | -Support Vector Machines (SVM) is a powerful, state-of-the-art algorithm based on linear and nonlinear regression |  |  |
| Advantages of SVM |  |  |  |  |  |  |  |  |  | -The advantage of the SVM is that, by use of the so-called ‘‘kernel trick’’, the distance between a molecule and the hyper plane can be calculated in a transformed (nonlinear) feature space, lacking of the explicit transformation of the original descriptors. |  |  |
| Disadvantages of SVM |  | -limitation of SVMs is the low speed of the training |  |  |  |  |  |  |  |  |  |  |
| Comparison | |  |  |  |  |  | | --- | --- | --- | --- | --- | | Name of Classifier | Evaluation Parameter | | | | | Accuracy | Precision | Recall | F-measure | | Logistic regression |  |  |  |  | | Naïve bayes |  |  |  |  | | Stochastic gradient descent |  |  |  |  | | K-nearest neighbor |  |  |  |  | | Decision trees |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | parameter | LR | BA | SGD | KNN | DT | RF | SVM | | Deterministic/non-deterministic |  | Non-deterministic |  | Non-deterministic | deterministic |  |  | | Optimal data size |  | big |  | Small | big |  |  | | speed |  | Faster than KNN |  | Slow on large data | fastest |  |  | | dataset |  | Good with noisy data |  | Poor with noisy data | Good with noisy data |  |  | | accuracy |  | High(only with high no. of records) |  | high | high |  |  | | Learning paradaigm |  | Bayesian network |  | Instance based | Inductive logic | Inductive logic |  | | | | | | | | | | | | |
| LR/BA |  |  |  |  |  |  |  |  |  |  |  |  |
| LR/SGD |  |  |  |  |  |  |  |  |  |  |  |  |
| LR/KNN |  |  |  |  |  |  |  |  |  |  |  |  |
| LR/DT |  |  |  |  |  |  |  |  |  |  |  |  |
| LR/RF |  |  |  |  |  |  |  |  |  |  |  |  |
| LR/SVM |  |  |  |  |  |  |  |  |  |  |  |  |
| BA/SGD |  |  |  |  |  |  |  |  |  |  |  |  |
| BA/KNN |  |  |  |  |  |  |  |  |  |  |  |  |
| BA/DT |  |  |  |  |  |  |  |  |  |  |  |  |
| BA/RF |  |  |  |  |  |  |  |  |  |  |  |  |
| BA/SVM |  |  |  |  |  |  |  |  |  |  |  |  |
| SGD/KNN |  |  |  |  |  |  |  |  |  |  |  |  |
| SGD/DT |  |  |  |  |  |  |  |  |  |  |  |  |
| SGD/RF |  |  |  |  |  |  |  |  |  |  |  |  |
| SGD/SVM |  |  |  |  |  |  |  |  |  |  |  |  |
| KNN/DT |  | KNN is more stable than descision trees. A lerning method is termed unstable if small changes in the training-test set split can result in large changes in the resulting classifier. |  |  |  |  |  |  |  |  |  |  |
| KNN/RF |  |  |  |  |  |  |  |  |  |  |  |  |
| KNN/SVM |  |  |  |  |  |  |  |  |  |  |  |  |
| DT/RF |  |  |  |  |  |  |  |  |  |  |  |  |
| DT/SVM |  |  |  |  |  |  |  |  |  |  |  |  |
| RF/LR |  |  |  |  |  |  |  |  |  |  |  |  |
| RF/SVM |  |  |  |  |  |  |  |  |  |  |  |  |