Machine learning algorithms

ASSIGN BUSTER

Clustering: TypeUnsupervised learning, class type clusterningClustering: DefinitionMethods to assign a set of objects into groups. Objects in a cluster are more similar to each other than to those in other clusters. Enables understanding of the differences as well as the similarities within the data. Clustering: Preference BiasPrefers data that is in groupings given some form of distance (Euclidean, Manhattan, or others)Clustering: Restriction BiasNo restrictionClustering: FlavorsWard hierarchical clustering, k-means, Gaussian Mixture Models, spectral, Birch, Affinity propogation, fuzzy clusteringClustering: K-MeansFor a given K, finds K clusters by iteratively moving cluster centers to the cluster centers of gravity and adjusting the cluster set assignments. K-Nearest Neighbors: TypeSupervised learning, instance basedK-Nearest Neighbors: DefinitionK-NN is an algorithm that can be used when you have a objects that have been classified or labeled and other similar objects that haven't been classified or labeled yet, and you want a way to automatically label them. K-Nearest Neighbors: Pros1: Simple; 2: Powerful; 3: Lazy, no training involved; 4: Naturally handles multiclass classification and regressionK-Nearest Neighbors: Cons1: Performs poorly on high-dimensionality datasets; 2: Expensive and slow to predict new instances; 3: Must define a meaningful distance function; K-Nearest Neighbors: Preference BiasGood for measuring distance based approximations, good for outlier detectionK-Nearest Neighbors: Restriction BiasLow-dimensional datasetsK-Nearest Neighbors: Example Applications1: Computer security: intrusion detection; 2: Fault detection in semiconducter manufacturing; 3: Video content retrieval; 4: Gene expressionDecision Trees: DefinitionEach node in the tree tests a single attribute, decisions are represented by the edges leading away from that node with leaf nodes representing the final decision. Decision Trees: Pros1: Fast; 2: Robust to noise and missing values; 3: AccurateDecision Trees: Cons1: Complex trees are hard to interpret; 2: Duplication within the same sub-tree is possibleDecision Trees: Best UsesDecision Trees: Restriction BiasDecision Trees: Example Applications1: Star classification; 2: Medical diagnosis; 3: Credit risk analysisDecision Trees: FlavorsCART, ID3Hidden Markov Models: TypeSupervised or unsupervised with class: MarkovianHidden Markov Models: DefinitionMarkov models are a kind of probabilistic model often used in language modeling. The observations are assumed to follow a Markov chain, where each observation is independent of all past observations given the previous one. Hidden Markov Models: ProsMarkov chains are useful models of many natural processes and the basis of powerful techniques in probabilistic inference and randomized algorithms. Hidden Markov Models: ConsHidden Markov Models: Preference BiasGenerally works well for system information where the Markov assumption holdsHidden Markov Models: Restriction BiasPrefers time series data and memoryless informationHidden Markov Models: Example ApplicationsTemporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, and bioinformatics. Hidden Markov Models: FlavorsMarkov chains, Hidden Markov ModelsLinear Regression: TypeSupervised learning, regression classLinear Regression: DefinitionTrying to fit a linear continuous function to the data to predict results. Can be univariate or multivariate. Linear Regression: Pros1: Very fast - runs in constant time, 2: Easy to understand the model, 3: Less prone to overfittingLinear Regression: Cons1: Unable to model complex relationships, 2: Unable to capture nonlinear relationships without first transforming the inputsLinear Regression: Preference Bias1: Prefers continuous variables; 2: A first look at a dataset; 3: Numerical data with lots of featuresLinear Regression: Restriction BiasLow restriction on problems it can solveLinear Regression: Example Applications1: Fitting a lineNaive Bayes: TypeSupervised learning; used for classification; probabalistic approachNaive Bayes: DefinitionGiven its simplicity and the assumption that the independent variables are statistically independent, Naive Bayes models are effective classification tools that are easy to use and interpret. Naive Bayes is particularly appropriate when the dimensionality of the independent space is high. For the reasons given above, Naive Bayes can often outperform other more sophisticated classification methods. A variety of methods exist for modeling the conditional distributions of the inputs including normal, lognormal, gamma, and Poisson. Naive Bayes: Pros1: Easy to use and interpret; 2: Works well with high dimensional problemsNaive Bayes: ConsNaive Bayes: Preference BiasWorks on problems where the inputs are independent from each otherNaive Bayes: Restriction BiasPrefers problems where the probability will always be greater than zero for each classNaive Bayes: Example ApplicationsNaive Bayes: FlavorsA variety of methods exist for modeling the conditional distributions of the inputs including normal, lognormal, gamma, and Poisson. Neural Networks: TypeSupervised learning; nonlinear functional approximationNeural Networks: DefinitionWith experience, networks can learn, as feedback strengthens or inhibits connections that produce certain results. Each layer depends on the calculations done on the layer before it. Neural Networks: Pros1: Extremely powerful, can model even very complex relationships;

2: No need to understand the underlying data;

3: Almost works by " magic" Neural Networks: Cons1: Prone to overfitting;

2: Long training time;

3: Requires significant computing power for large datasets;

4: Model is essentially unreadable;

5: Work best with " homogenous" data where features all have similar meaningsNeural Networks: Preference BiasPrefers binary inputsNeural Networks: Restriction BiasLittle restriction biasNeural Networks: Example Applications1: Images; 2: Video; 3: " Human-intelligence" type tasks like driving or flying; 4: RoboticsNeural Networks: FlavorsDeep learningSupport Vector Machines: TypeSupervised learning for defining a decision boundarySupport Vector Machines: DefinitionDivides an instance space by finding the line that is as far as possible from both classes. This line is called the " maximum-margin hyperplane". Only the points near the hyperplane are important. These points near the boundary are called the support vectors. Support Vector Machines: Pros1: Can model complex, nonlinear relationships; 2: Robust to noise (because they maximize margins)Support Vector Machines: Cons1: Need to select a good kernel function; 2: Model parameters are difficult to interpret; 3: Sometimes numerical stability problems; 4: Requires significant memory and processing powerSupport Vector Machines: Preference BiasWorks where there is a definite distinction between two classificationsSupport Vector Machines: Restriction BiasPrefers binary classification problemsSupport Vector Machines: Example Applications1: Text classification; 2: Image classification; 3: Handwriting recognition