Data Mining🡪non-trivial extraction of implicit, previously unknown and potentially useful information from data. KDD= KnowledgeDiscoveryofData. knowledge is useful information, esoteric, and useful in a practical sense. we are reading between the lines making the implicit into explicit and read between the lines. supervised learning: there is a training set. unsupervised learning: no training set. ex: clustering. Lazy lernr🡪no model, learner looks at data as data arrives then compute prediction. Rote classifier: memorizes entire training data and performs classification only if attributes of a test instance match a training instance exactly.||Eager lernr🡪new data fed into model, model output’s prediction, eager lernr does bulk of work once. Classifier🡪algor. that classifies.

prediction methods: (classification) use some variables to predict unknown or future values of other variables descriptive methods: (clustering) find human-interpretable patterns that describe data. defining categories into new/better categories.

bayesian belief networks: naïve bayes classifiers too rigid for problems in which attributes are somewhat correlated. BBN is graphical representation of probabilistic relationships among a set of random variables. Two parts of BBN: 1. a directed acyclic graph DAG 2. a probability table associating each node to its immediate parent nodes. *a node in a BBN is conditionally independent of its non-descendants if its parents are known.* Constructing the network requires a lot of costs; BBN= capturing prior knowledge of a domain using graphical model, well suited for dealing with incomplete data, data is combined probabilistically with prior knowledge, BBN is robust to model overfitting.

ensemble methods: improve classification accuracy by aggregating the predictions of multiple classifiers. ensemble meth. constructs base classifiers from training data and performs classification by taking a vote on the predications made by each classifier. ensemble PERFORMS BETTER THAN SINGLE DECISION TREE CLASSIFIER. two conditions for ensemble to (work) perform better than single classifier: 1. the base classifiers should be independent of each other (must be diverse) 2. base classifiers should do better than a classifier that performs random guessing (classifier must be accurate). difficult to assure total independence. construct multiple classifiers form original data then aggregate their predictions. ensemble classifiers can be constructed 1. by manipulating the training set: create multiple training sets by resampling the original data according to some sampling distribution. BAGGING and BOOSTING are two examples of ensemble methods. 2. by manipulating their input features: subset of input attributes chosen to form each training set, that subset chosen randomly or by domain experts. RANDOM FOREST is ensemble method that manipulates input features and uses decision trees as its base classifiers. 3. by manipulating class labels: # classes is large. transform data into a binary class problem by randomly partitioning the class labels into two disjoint subsets. error correcting output coding is example of this. 4. manipulating the learning algor: manipulate the learning model algorithm to generate different models. bagging: repeatedly samples w replacement from a data set according to uniform probability distribution. sampling done w replacement, so some instances may appear several times whiles other instances omitted. on average bootstrap sample d contains on average 63% original training data bc each sample has probability 1-(1-1/N)N of being selected in d. large N makes previous equation converge to 63%. bagging improves generalization error by reducing variance of base classifiers. bagging is less susceptible to model overfitting when applied to noisy data. boosting boosting is an iterative procedure used to adaptively change the distribution of training examples so that the base classifiers will focus on examples that are hard to classify. boosting assigns a weight to each training example & may change to weight at end of each boosting round. weights can be 1. used as sampling distribution to draw set of bootstrap samples from original data. 2. used by base classifier to learn a model that is biased toward higher-weight examples. boosting is sequential ensemble, aim to decrease bias. random forests: class of ensemble methods specifically designed for decision tree classifiers, combines predictions made by multiple decision trees, where each tree is generated based on values of an independent set of random vectors. rand. vects. are generated from a fixed probability distribution where the probability distribution is varied to focus on examples that are hard to classify. random subspace method: an ensemble classifier that consists of several classifiers trained on the same dataset using only a (random) subset of the available features, a generalization of random forest, can be composed from any underlying classifiers.

k-NN classification: changing k value may change classification, training set has high impact on test case, k too small leads to biased variance & k too small leads to OVERFITTING because noise more pronounced in small k, higher k means less variance & too large of a k has points too far out influencing test point.

underfitting: both the training errors and test errors are too large, happens when developed model is made very simple. overfitting: problem arises when training errors are small, but test errors are large, results in decision trees that more complex than necessary, training error no longer provides good estimate of how well the tree will perform on unseen records. The model doesn’t generalize well from training data to unseen data. pruning: tree pruning addresses overfitting, tree pruning is the process of adjusting decision tree to minimize misclassification error done in pre prun or post prun. prepruning: halting subtree construction at some node after checking some measures (information gain/gini index. etc), may stop growth process prematurely, faster than postpruning bc doesn’t wait for full construction of decision tree. postpruning: grow decision tree in its entirety, trim nodes in bottom-up fashion replacing a node with a leaf, preferable to prepruning bc of interaction effect- effects which arise after interaction of several attributes. prepruning suppresses growth by evaluating each attribute individually so overlooks effects due to interaction of several attributes.

more complex v less complex model: no, k-NN one of the simplest and best models for tasks like recommendations. more complex model costs more time and money also!

