# Question we can have it predict future

Business, Decision Making

Question 1Give a definition of a learning system. What are the basiccomponents at the learning system? There are three different tasks in machinelearning (ML): supervised learning, unsupervised learning, and reinforcementlearning.

How are they different? What kind of data is provided for eachlearning task? How is reinforcement learning (RL) related to supervisedlearning? How is RL related to unsupervised learning? Explain how to teach achess—playing program using supervised, unsupervised, and reinforcementlearning paradigms, respectively. Learning is anypractice by which a system improves Performance from experience. For example, let take school as an example. School have teacher, books and other resources. Students they have they get experience from their teacher, books and otherresources and as a result of this they improve their performance. Basic components I. Performance II.

Task(future task) III. Trainingexperience (datasets)Supervise means to observe anddirect the execution of a task. Means supervising a machine learning model thatmight be able to produce classification regions. Teach the model by training itwith some data from a labeled dataset than load the model with knowledge sothat we can have it predict future instances. Generally speaking, the model istrained on a labeled dataset, so it can predict the outcome of out of sampledata. There is 2 type of supervised learning: classification and regression. Unsupervised learning is exactly asit sounds, let the model work on its own to discover information that may not bevisible for our eyes. It uses machine learning algorithms that extractconclusions on unlabeled data.

Unsupervised learning has more difficultalgorithms than supervised learning, since we know little to no informationabout the data, or the outcomes that are to be expected. With unsupervisedlearning, we’re looking to find things such as group, clusters perform densityestimation and dimensionality reduction. In supervised learning, however, weknow what kind of data we’re dealing with, since it is labelled data. A system interacts with environmentto perform some task in return of this environment give some reward to thesystem reward is either positive or negative on the base of this reward systemimprove his performance and again perform task until system achieve maximumpositive reward. Reinforcement learning is something between supervised andunsupervised learning it told us when we are wrong through negative reward butit didn’t tell us which way to get maximum positive reward it should exploreall the possibility. In comparison to supervisedlearning, unsupervised learning has: fewer tests and fewer model that can beused in order to ensure the outcome of the model is accurate. As suchunsupervised learning create a less controllable environment, as the machine iscreating outcomes for us.

The biggest difference between supervised andunsupervised learning is that supervised learning deals with labeled data whileunsupervised learning deals with unlabeled data. In supervised learning, wehave machine learning algorithms for classification, and regression. Classification is the organization of labeled data and regression is theprediction of trends in labeled data to determine future outcomes. In unsupervisedlearning, we have clustering. Clustering is the analysis of patterns andgroupings of unlabeled data. Reinforcement learning is not exactly supervisedlearning because it doesn’t rely strictly on labeled data. It actually relieson “ reward”. But it’s not unsupervised learning either, since we know straightwhen we model our “ learner” which is the estimated reward.

Supervised learning: Labeled Data. Unsupervised learning: Unlabeled Data. Reinforcement learning: Have no data weconstruct a model that generates data based on reward.

Question 2Consider-solving the problem ofcharacter—image classification using a back-propagation learning neuralnetwork. What is the model structure? What training data should be provided? How does the neural network (model) learn from data? Explain the procedure forlearning. Question 3What is generalization? We’ve mostly been talking about thetraining data right so and classifiers and the algorithms use the training datato build the predictors so the training data attributes X along with some targetsY which could be a clear at a class or a number or something so that’s what weused to train to build our model but then the reason we’re building thispredictor is that sometime tomorrow or in the future we’re going to get newdata and on that data.

What is over? tting in learning? Overfitting is happened when you find or when you learn a predictorthat fits the training data a little bit to wealth so it’s usually happens whenyour predictor the function that you’re predicting is complex enough andflexible enough to fit any kind of sort of noise in the training data so thoseare patters that are present in the training data that will not be presenttomorrow in the future data that you see so the when that happens you say youover fit. Explain the over? tting phenomena using the polynomial interpolationproblem. How can we avoid the over? tting problem in learning? Debugging and diagnosing things they can go wrong with learningalgorithms will give you specific tool to recognize when overfitting. If wethink overfitting is occurring what can we do to address when we had 1 or 2dimensional data so we could just plot the hypothesis and see what was going onand select the appropriate degree polynomial. We could just plot hypothesis andif it was fitting the sort of very wiggly function that goes all over the placeto predict and we could then use appropriate degree polynomial. So plotting thehypothesis could be one way to try to decide what degree polynomial to use butthat does always work. In fact it when we have so many features it also becomesmuch harder to plot the data and becomes much harder to visualize it to decidewhat features to keep or not so concretely suppose if  we’re trying to predict housing pricessometimes we can just have a lot of different features and all of thesefeatures seem you know maybe they seem kind of useful but if we have a lot of features and very little trainingdata then overfitting can become a problem in order to address overfittingthere are two main options for things that we can do the first option is to tryto reduce the number of features concretely one thing we could do is manuallylook through the list of features and use that to try to decide which are themore important features and therefore which are the features we should keep andwhich other features we should throw out there are algorithms for automaticallydeciding which features the key and which features to throw out this idea ofreducing the number of features can work well and can reduce overfitting andwhen we talk about model selection we’ll go into this in much greater depth butthe disadvantage is that by throwing away some of the features is also throwingaway some of the information you have about the problem for example maybe all ofthose features are actually useful for predicting the price of a house so maybewe don’t actually want to throw some of our information or throw some of ourfeatures away.

Regularization: we’re going to keep all the features but we’regoing to reduce the magnitude this method works well we’ll see when we have alot of features each of which contributes a little bit to predicting the value. How can we avoid the overfitting problem in back-propagation neuralnetworks? DownhillIf you have very uneven training andtest data structure, try to fix it. E. g.

the share of classes zero and one inboth datasets should be equal. You can also randomly drop some neurons during training. For randomly stopping neurons u can use the l2 loss function Question 4Give a list of machine learning models for supervised learning. Howare they different? What are their similarities? What representations do theyuse? What ML methods use tree structures for representing their model? Whatmethods use graph or network representations? What methods use list structuresor rule sets for representing models? o SupportVector Machineso linearregressiono logisticregressiono naiveBayeso lineardiscriminant analysiso decisiontreeso k-nearestneighbor algorithmo NeuralNetworks (Multilayer perceptron). Classification (1R, Naive Bayes, Decision tree learning algorithm such as ID3 CART and so on)Numeric Value PredictionDecision tree continuous and categorical inputs. While decision trees classifyquickly, the time for building a tree may be higher than another type ofclassifier Decision trees suffer from a problem of errors propagatingthroughout a tree Decision trees can be used to help predict the future Thetrees are easy to understand Decision trees work more efficiently with discreteattributes The trees may suffer from error propagationSVM continuous value inputsNaïve Bayes A simple but effective learningsystem. Each piece of data that is to be classified consists of a set ofattributes, each of which can take on a number of possible values. The data arethen classified into a single classification.

Advantages: –Fast to train(single scan). Fast to classify–Notsensitive to irrelevant features –Handles real and discrete data –Handlesstreaming data well Disadvantages: –Assumes independence of features KNN continuous value inputs o DecisionTrees are fast to train and easy toevaluate and interrupt. o Supportvector machine gives good accuracy, power offlexibility from kernels. o Neuralnetwork is slow to converge and hard to setparameters but if done with care it works wellso Bayesianclassifiers are easy to understand. Question 5Machine learning methods can be de? ned by three dimensions: type oflearning data, model structure, and learning algorithm. Describe these aspectsfor each of the following methods.

Linear regression Typeof learning data Continuous and categorical inputsModel structureLinearregression is a very simple approach for supervised learning. Though it mayseem somewhat dull compared to some of the more modern algorithms, linearregression is still a useful and widely used statistical learning method. Linear regression is used to predict a quantitative response Y from thepredictor variable X. Linear Regression is made with anassumption that there’s a linear relationship between X and Y. Y = W0 + W1X, where X is the explanatoryvariable and Y is the dependent variable. The slope of the line is W1, and W0 is the intercept (the value of y when x= 0)Learningalgorithm Multi-layer preceptors, Support vectormachine, Random forestDecision trees Typeof learning data Continuous and categorical inputsModel structureA tree has manyanalogies in real life, and turns out that it has influenced a wide area ofmachine learning, covering both classification and regression. In decisionanalysis, a decision tree can be used to visually and explicitly representdecisions and decision making.

As the name goes, it uses a tree-like model ofdecisions. Though a commonly used tool in data mining for deriving a strategyto reach a particular goal, it’s also widely used in machine learning, whichwill be the main focus of this article. The goal of Decision Tree is to createa model that predicts the value of a target variable by learning simpledecision rules inferred from the data featuresLearningalgorithm Random forest, C4. 5, ID3, C5. 0 etc. Neural networks Typeof learning data Continuous and categorical inputsModel structureThe dendritescarry the signal to the cell body where they all get summed. If the final sumis above a certain threshold, the neuron can fire, sending a spike along itsaxon. In the computational model, we assume that the precise timings of thespikes do not matter, and that only the frequency of the firing communicatesinformation.

we model the firing rate of the neuron with an activation function(e. g. sigmoid function)Learningalgorithm K-means clustering Typeof learning data Continuous inputsModel structureK-meansclustering is a type of unsupervised learning, which is used when you haveunlabeled data (i. e., data without defined categories or groups). The goal ofthis algorithm is to find groups in the data, with the number of groupsrepresented by the variable K. The algorithm works iteratively to assign eachdata point to one of K groups based on the features that are provided.

Datapoints are clustered based on feature similarity. Rather thandefining groups before looking at the data, clustering allows you to find andanalyze the groups that have formed organically. The “ Choosing K” section below describes how the number of groups can be determined. Each centroidof a cluster is a collection of feature values which define the resultinggroups.

Examining the centroid feature weights can be used to qualitativelyinterpret what kind of group each cluster represents. Pizza example which you told us inthe class population divided according to 3 branches of pizza. LearningalgorithmNaive Bayes Classi? er Typeof learning dataModel structureThis lets us examine the probability of anevent based on the prior knowledge of any event that related to the formerevent. So for example, the probability that price of a house is high, can bebetter assessed if we know the facilities around it, compared to the assessmentmade without the knowledge of location of the house.

Bayes’ theorem doesexactly that. Aboveequation gives the basic representation of the Bayes’ theorem. Here A and B aretwo events and, P(A|

B): the conditional probability that event A occurs, given that B has occurred. This is also known as the posterior probability. P(A)and P(B): probability of A and B without regard of each other.

P(B| A): the conditional probability that event B occurs, given that A has occurred. Learningalgorithm Question 6Explain how machine learning can be used for the followingapplications. You may refer to existing work and discuss it. Optical character recognitionPredicting customer’s response tocoupon mailsSpeech recognitionAutonomous car drivingClustering the types of customers ofinternet shopping mallsBoard game playing (e. g., Chess, backgammon)Helicopter control