A look at

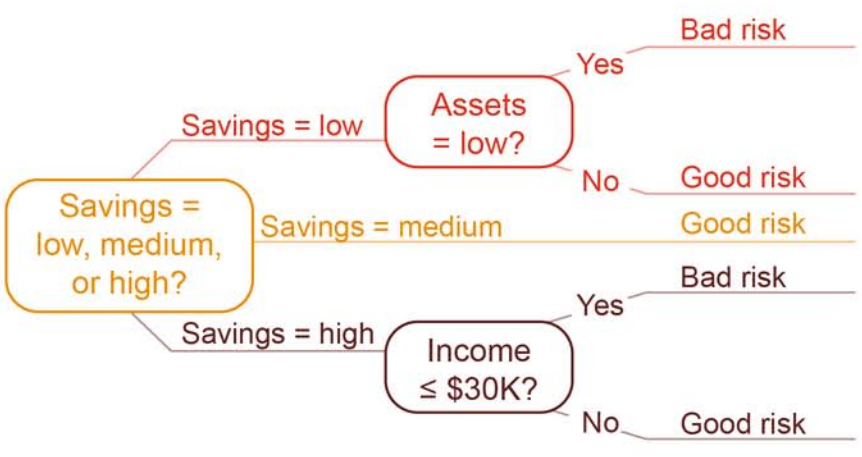
Machine learning methods

Introduction

Which machine learning algorithm should you use? A lot depends on the characteristics and the amount of the available data, as well as your training goals, in each particular use case. Avoid using the most complicated algorithms unless the end justifies more expensive means and resources. Here are some of the more common algorithms ranked by ease of use.

Decision trees

Decision tree analysis typically uses a hierarchy of variables or decision nodes that, when answered step by step, can classify a given customer as creditworthy or not, for example.

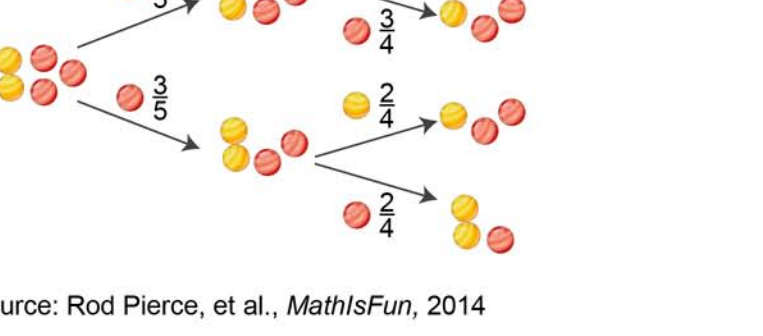


Advantages	Use cases
Decision trees are useful when evaluating lists of distinct features, qualities, or characteristics of people, places, or things.	Rule-based credit risk assessment, horse race performance prediction

Source: Daniel T. Larose and Chantal D. Larose, *Data Mining and Predictive Analytics,* 2nd Edition, John Wiley & Sons, 2015

Support vector machines

Support vector machines classify groups of data with the help of hyperplanes.

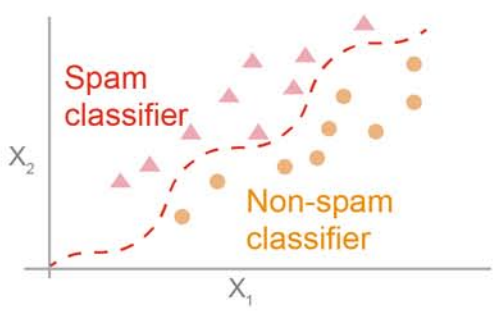


Advantages	Use cases
Support vector machines are good for the binary classification of X versus other variables and are useful whether or not the relationship between variables is linear.	News categorization, handwriting recognition

Source: Matthew Kelly, *Computer Science: Source,* 2010

Regression

Regression maps the behavior of a dependent variable relative to one or more dependent variables. In this example, logistic regression separates spam from non-spam text.



Advantages	Use cases
Regression is useful for identifying continuous (not necessarily distinct) relationships between variables.	Traffic flow analysis, email filtering

Naive Bayes classification

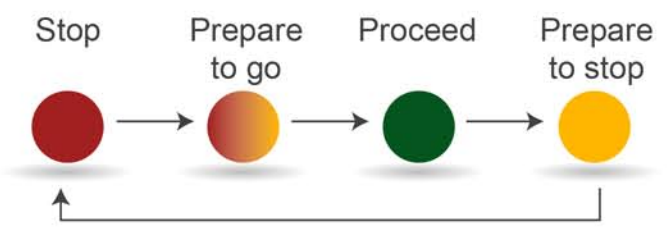
Naive Bayes classifiers compute probabilities, given tree branches of possible conditions. Each individual feature is "naive" or conditionally independent of, and therefore does not influence, the others. For example, what's the probability you would draw two yellow marbles in a row, given a jar of five yellow and red marbles total? The probability, following the topmost branch of two yellow in a row, is one in ten. Naive Bayes classifiers compute the combined, conditional probabilities of multiple attributes.

Advantages	Use cases
Naive Bayes methods allow the quick classification of relevant items in small data sets that have distinct features.	Sentiment analysis, consumer segmentation

Source: Rod Pierce, et al., *MathIsFun,* 2014

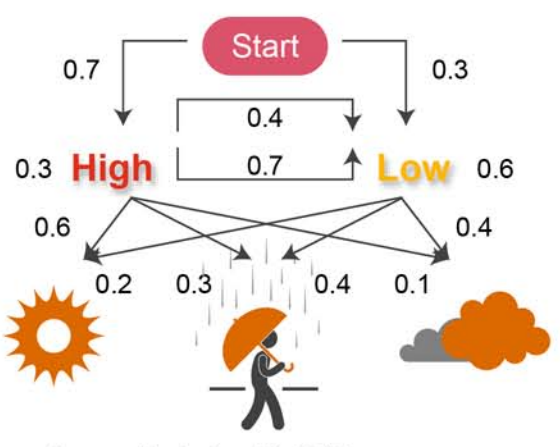
Hidden Markov models

Observable Markov processes are purely deterministic—one given state always follows another given state. Traffic light patterns are an example.



Source: Derek Kane, 2015

Hidden Markov models, by contrast, compute the probability of hidden states occurring by analyzing observable data, and then estimating the likely pattern of future observation with the help of the hidden state analysis. In this example, the probability of high or low pressure (the hidden state) is used to predict the likelihood of sunny, rainy, or cloudy weather.

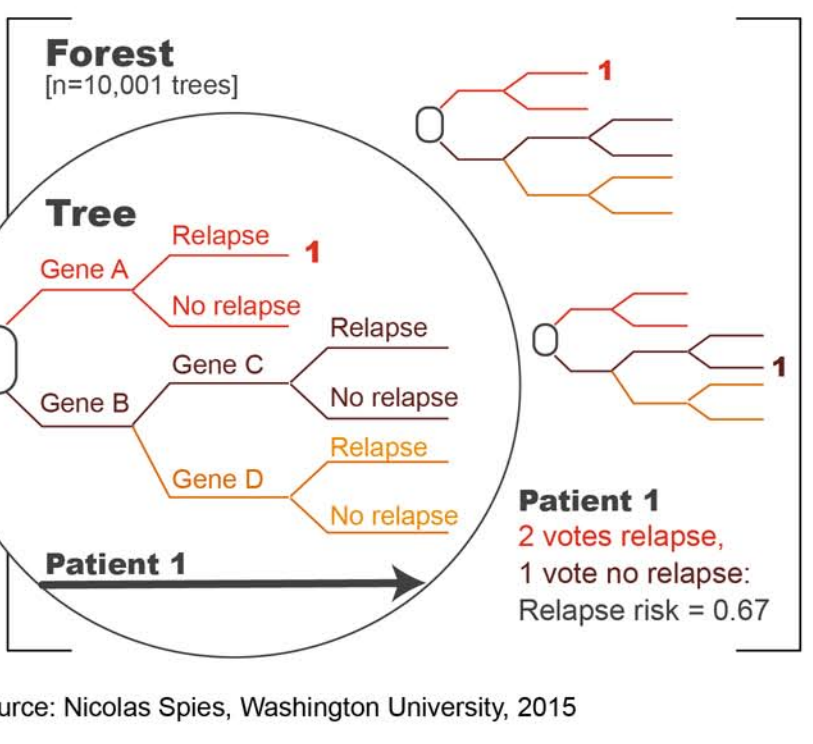


Advantages	Use cases
Tolerates data variability and effective for recognition and prediction.	Facial expression analysis, weather prediction

Source: Leonardo Guizzetti, 2012

Random forest

Random forest algorithms improve the accuracy of decision trees by using multiple trees with randomly selected subsets of data. This example reviews the expression levels of various genes associated with breast cancer relapse and computes a relapse risk.



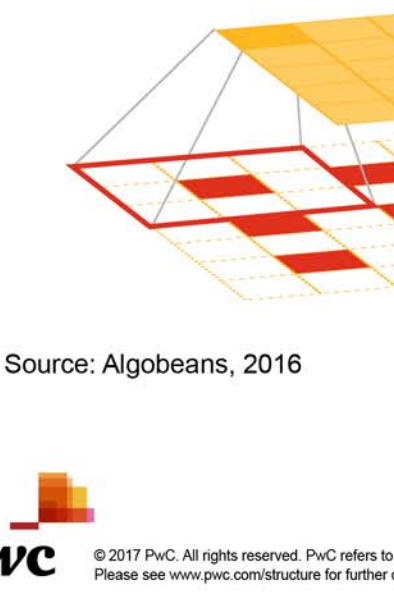
Advantages	Use cases
Random forest methods prove useful with large data sets and items that have numerous and sometimes irrelevant features.	Customer churn analysis, risk assessment

Source: Nicolas Spies, Washington University, 2015

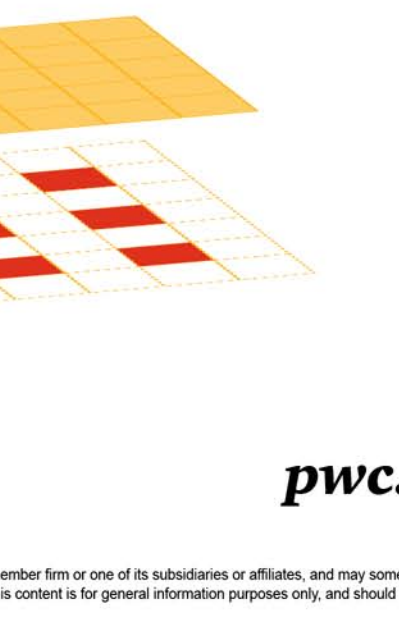
Recurrent neural networks

Each neuron in any neural network converts many inputs into single outputs via one or more hidden layers. Recurrent neural networks [RNNs] additionally pass values from step to step, making step-by-step learning possible. In other words, RNNs have a form of memory, allowing previous outputs to affect subsequent inputs.

Non-recurrent feed-forward neural network



Recurrent neural network—includes loops



Advantages	Use cases
Recurrent neural networks have predictive power when used with large amounts of sequenced information.	Image classification and captioning, political sentiment analysis

Source: Joseph Wilks, 2012

Long short-term memory & gated recurrent unit neural networks

Older forms of RNNs can be lossy. While these older recurrent neural networks only allow small amounts of older information to persist, newer long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks have both long- and short-term memory. In other words, these newer RNNs have greater memory control, allowing previous values to persist or to be reset as necessary for many sequences of steps, avoiding "gradient decay" or eventual degradation of the values passed from step to step. LSTM and GRU networks make this memory control possible with memory blocks and structures called gates that pass or reset values as appropriate.

Source: Genevieve Orr, et al., Williamette University, 1999

Advantages	Use cases
Long short-term memory and gated recurrent unit neural networks have the same advantages as other recurrent neural networks and are more frequently used than other recurrent neural networks because of their greater memory capabilities.	Natural language processing, translation

Convolutional neural networks

Convolutions are blends of weights from a subsequent layer that are used to label the output layer.

Advantages	Use cases
Convolutional neural networks are most useful with very large data sets, large numbers of features, and complex classification tasks.	Image recognition, text to speech, drug discovery

Source: Algobeans, 2016

pwc.com/NextinTech