An Overview of Machine Learning

What is Machine Learning?

Machine learning is the study of developing and utilizing algorithms which are able to make predictions and decisions without explicitly being programmed by a human.

The “machine” learns from sample data called training data, constructing a mathematical model to describe the data.

A wide field with dozens of algorithms and methods suitable for different data structures and tasks.

Applications of Machine Learning

- Computer Science
  - Data mining, pattern recognition, data quality, computer vision
- Sciences
  - Bioinformatics, medical diagnosis, cheminformatics, mathematics
- Financial
  - Economics, banking, insurance, stock market predictions
- Social
  - Social media, networking, speech recognition, search engines

Essentially anything involving data!

Types of Learning

Supervised

- Training data made up of labeled input-output pairs, algorithm creates function to produce output for an input

Unsupervised

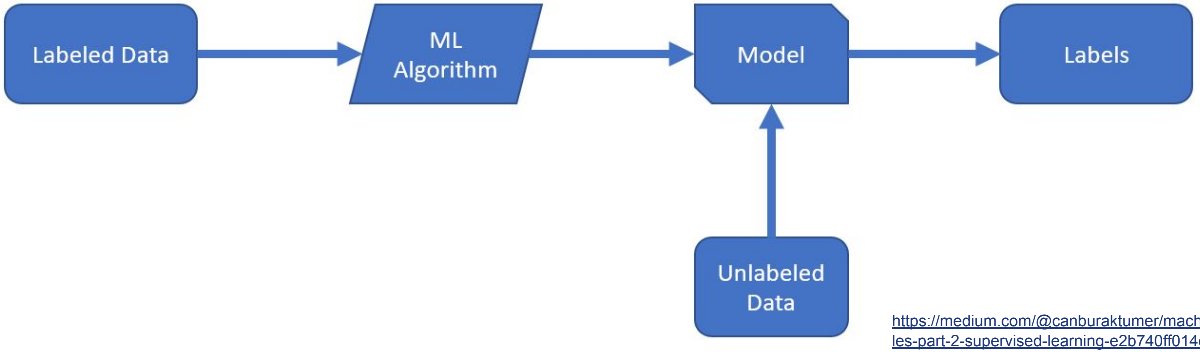
- Training data doesn’t have pre-existing labels, helps find previously unknown patterns

Reinforcement

- Reward based, works on the principle of feedback using training data

Supervised Learning

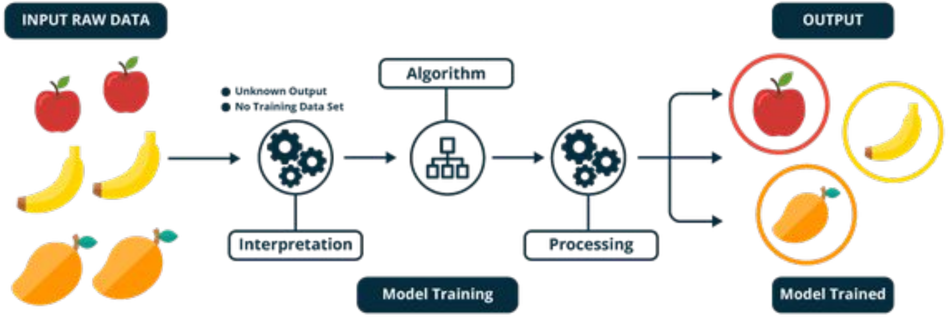
- Works with training data made up of labeled input-output pairs
- Analyzes the patterns found in the training data and constructs a resulting function which can be used to map new examples
- Often work by optimizing a loss function given a set of inputs and outputs, so given a new input it can give an accurate output
- Can be divided into two categories, generative, using a joint probability distribution, and discriminative, using a conditional probability distribution



https://medium.com/@canburaktumer/machine-learning-basics-with-examples-part-2-supervised-learning-e2b740ff014c

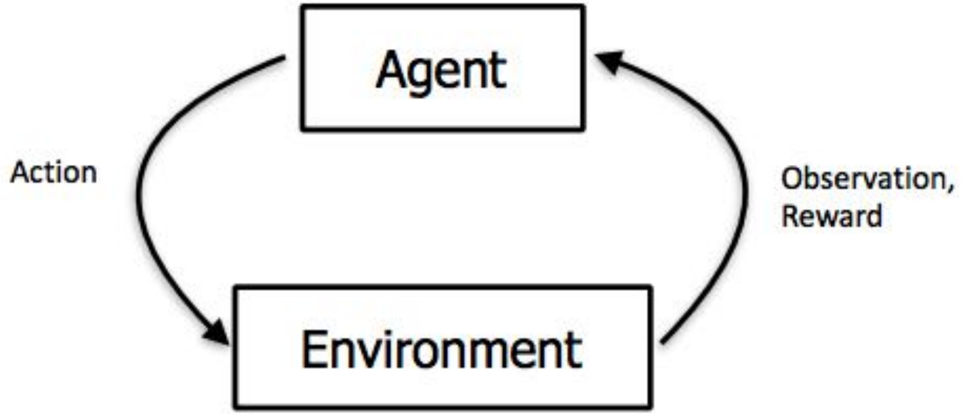
Unsupervised Learning

- Works with training data that has no pre-existing labels
- It is the algorithm's job to learn from the training data, find previously unknown patterns and perhaps assign labels.
- Model is constructed by discovering structures in the training data, extracting general rules
- Can be broken up into two categories, clustering, where data is grouped by similarity, and dimension reduction, where the most relevant features are found



Reinforcement Learning

- Is reward based learning, working on the principle of feedback with its training data, looking to find the best model to minimize an undesirable quantity or maximize a desirable quantity.
- It's training data does need to be labeled, as the model learns from experience.
- Works primarily through trial and error, and can be categorized as Markovian or evolutionary.



Types of Algorithms

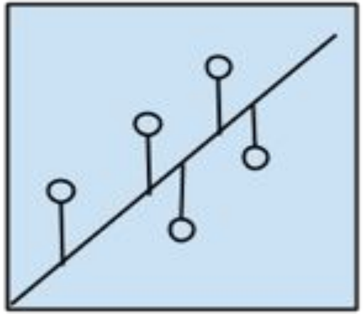
Different types of problems require different types of machine learning algorithms.

Performance of an algorithm can be evaluated based on the accuracy and insightfulness of the results, as well as how well the structure of the model matches the data.

Regression

Used to predict a continuous value, where given an input a continuous output can be found. The model finds the relationship between a dependent variable and an independent variable.

Includes linear, polynomial, support vector, decision tree, random forest, ordinary least squares, and logistic regressions.

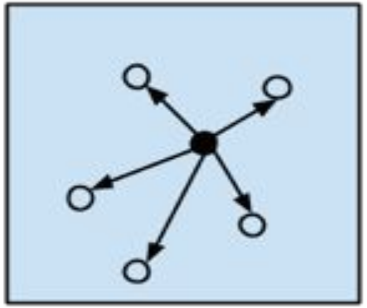


Regression Algorithms

Instance Based

Used to predict an output given an input. This type of learning doesn't create a model, rather it remembers instances in the training data and compares the current input to the instances in the training data.

Instance based learning needs to use algorithms to come to a conclusion such as k-Nearest Neighbor, Learning Vector Quantization, Support Vector machines.

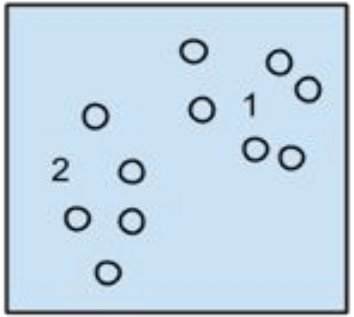


Instance-based Algorithms

Clustering

Works to organize data into groups who share common features. An ideal model has learned how to put together groups who have like features and far from groups with different features.

Algorithms to facilitate clustering include k-Means, k-Medians, and Hierarchical Clustering.

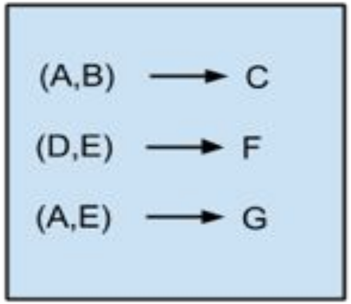


Clustering Algorithms

Association Rule

Association rule algorithms look to discover meaningful relationships between variables in data sets or rules which explain them. For example, it might find a rule for shopping if someone buys beef and hamburger buns they are likely to also buy cheese.

Algorithms to use Association Rule machine learning include Apiori, Eclat, and frequent pattern growth algorithms.



Association Rule Learning Algorithms

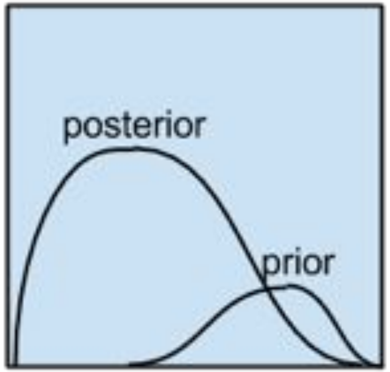
Bayesian

Bayesian machine learning tackles problems like regression and classification by utilizing Bayes' theorem. It allows us to calculate the probability of an event given knowledge of previous events.

Algorithms to perform Bayesian machine learning include Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes among others.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

https://towardsdatascience.com/what-is-bayes-rule-bb6598d8a2fd

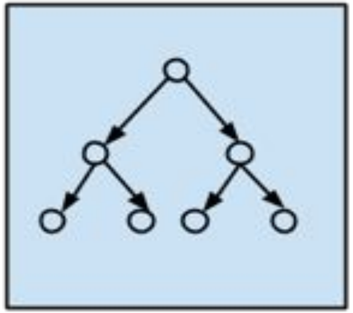


Bayesian Algorithms

Decision Tree

Decision tree algorithms construct a model of decisions based on the training data to come to a conclusion about the input data, such as classifying it to a category.

Popular algorithms include Classification and Regression Tree, Iterative Dichotomiser 3, Chi-Squared, and Decision Stump.

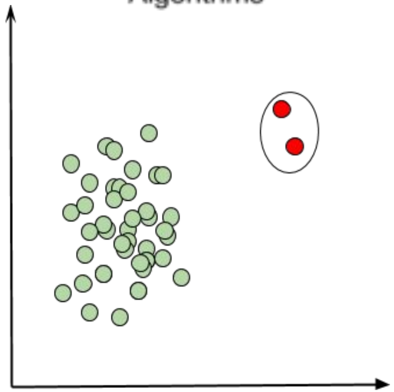


Decision Tree Algorithms

Anomaly Detection

Anomaly detection learns from the training data and looks to identify data points which do not conform to expected patterns or are vastly different than the other members in the data set.

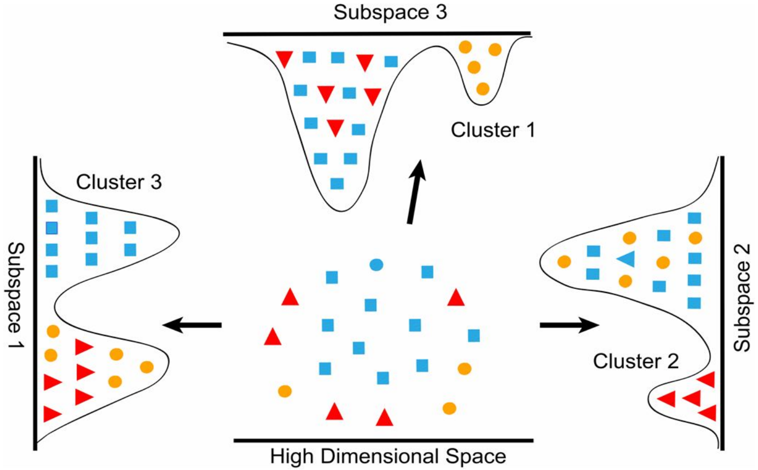
Anomaly detection algorithms can use supervised or unsupervised learning, and can implement neural networks, support vector machine learning, k-nearest neighbors, Bayesian networks, and decision trees.



https://blog.floydhub.com/introduction-to-anomaly-detection-in-python/

Dimensionality Reduction

Dimensionality Reduction is used to find the most relevant features of a data set needed to describe it, or to build a model relating to it. Dimensionality Reduction helps to avoid overfitting and can greatly reduce computational costs. Dimensionality Reduction is often applied to a data set, whose output is used in another Machine Learning algorithm. Dimensionality Reduction algorithms include Principal Component Analysis, Principal Component Regression, Partial Least Squares Regression, Sammon Mapping.

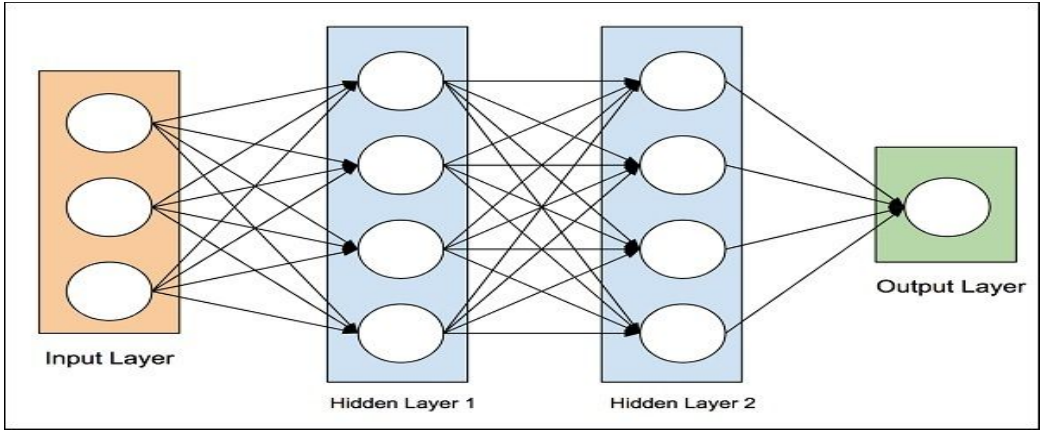


https://datascience.stackexchange.com/questions/130/what-is-dimensionality-reduction-what-is-the-difference-between-feature-selecti

Neural Network

Neural Network methods are inspired by biological neurons in the way different nodes in the system interact. The model learns from the training data to perform a task by looking at examples without being explicitly programmed a rule.

A Neural Network is composed of layers of “neurons” including an input layer, hidden layers, and an output layer. Information is passed from the input layer to the output layer through channels as it propagates through the hidden layers. Each channel has a weight which modifies the information from layer to layer until an output is reached. During training, the model’s output is compared with the true output and the model is improved based on the results.

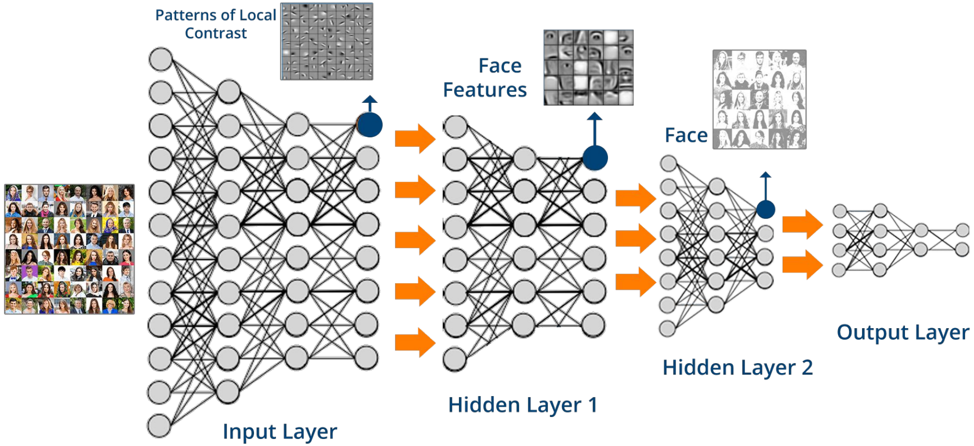


https://www.tutorialspoint.com/machine_learning/machine_learning_artificial_neural_networks.htm

Deep Learning

Deep Learning involves using neural networks on a large scale to tackle machine learning problems. Deep Learning algorithms are often used to model and analyze very large data sets, so they require a larger and more complex neural network. Deep Learning is often applied to visual data such as text and images as well as audio and video.

Some common Deep Learning algorithms are Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory Networks, and Deep Boltzmann Machine .



https://www.datacamp.com/community/tutorials/introduction-deep-learning

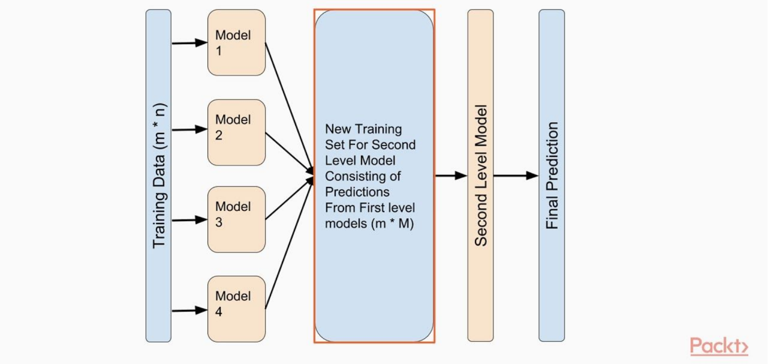
Kernel Methods

Kernel methods are algorithms which are used for pattern analysis. They work to provide a way to analyze data in a higher dimensional space by operating on it in its original space.

Are a key component in Support Vector Machines, and in forms of PCA, CCA and regression.

Ensemble

Ensemble Deep Learning models are made up of multiple Deep Learning models which are trained on their own and whose outputs are combined to come to a final conclusion.



https://www.youtube.com/watch?v=DCrcoh7cMHU

Choosing a ML method

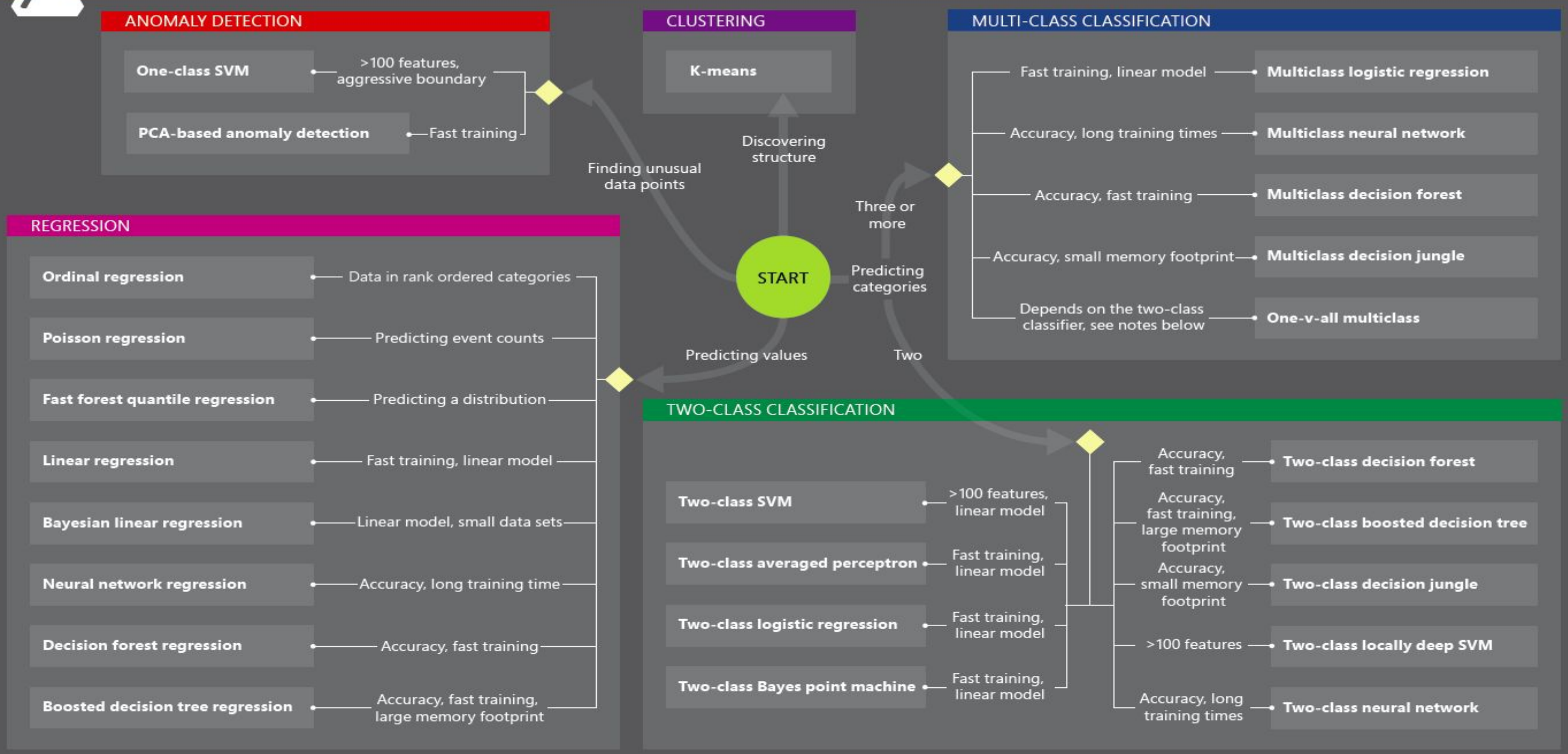
- Choosing the correct method requires:
    - Understanding the data
    - Understanding what the solution should look like
    - Fundamentals of ML algorithms.

- Choosing the correct method affects:
    - Accuracy
    - Insightfulness
    - Computation Time
    - Ease of Construction

Putting thought into which method you’d like to employ is worth it!

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com
Download this poster: http://aka.ms/MLCheatSheet

Microsoft

Useful Attributes/Definitions

There are some key attributes used in the cheat sheet worth explaining, which help the decision making process:

- Features refer to attributes or variables in the data which are provided for each data point. For example, if you are looking at a data set concerning humans, features of the data set might be Name, Age, Gender, Height, etc. The more variables in the data set, the larger the data set, given the same number of data points.
- Training speed or “fast training” refers to how long it is expected to teach the model, using training data, till the point the model can produce meaningful, accurate output.
- Accuracy refers to how well the output of the model matches the correct output.
- Classes are groups, labels, or categories we hope to put the data into.
- A linear model uses a linear mathematical function to describe the data, ie y=mx+b

Determining Your Objective

To choose the correct algorithm to use you need to identify what your objective is. Your objective might be:

- Predicting values, where given an input value you hope to find an accurate output value. The best way to do this is through regression.

- Predicting categories, where given inputs you hope to find meaningful labels to assign them. This can be split into two broader families, two-class classification and multi-class classification.

- Discovering structure, where you hope to find insightful ways to organize the input data or data set into groups.

- Anomaly detection, where the aim is to find unusual data points that are outliers or stand out in one way or another from the rest of the data set.

Predicting Values - Regression

Predicting values is best done by regression and can be implemented in many ways.

- Linear regression is a relatively simple solution and is well suited for simple problems where a linear model can be applied. For small data sets, Bayesian linear regression is a good solution.
- Neural networks and decision forests are good for more advanced regression problems that have a stronger sensitivity to accuracy. Neural networks involve a longer training time, while a decision forest will train faster. Boosted decision trees are more powerful, but require a larger memory footprint.
- Poisson regression is useful when trying to model counting and work with counting data.
- Fast forest quantile regression is helpful in modeling the distribution of a predicted value, such as gas prices, or performance of students on a test.
- Ordinal regression is great when your objective is somewhere in between regression and classification, where your input may be a value and the output is more of a rank order or category rather than strictly continuous

Predicting Categories - Two-Class Classification

When predicting how to put input into categories, and there are two possible categories, two-class classification methods are used.

- When looking for a simple, quick two-class algorithm, averaged perceptron, logistic regression, point machine, and for especially small data sets, SVM can be used. These models are relatively straightforward, linear models with fast training speeds.

- Decision tree, forest, and jungle two-class classification algorithms are best used for more complex two-class classification problems, and lead to higher accuracy.

- Two-class locally deep SVM classifiers are especially suitable for larger data sets containing roughly, or more than, 100 features.

- Two-class neural network classifiers have high accuracy, but require a longer time to train compared to other models.

Predicting Categories - Multi-Class Classification

When predicting how to put input into categories, and there are two possible categories, two-class classification methods are used.

- Multitask logistic regression can be used on simple multi-class classification problems. It uses a linear model which leads to fast training speeds.

- Similar to two-class classification, neural network, decision forest, and decision jungle algorithms can be used for more complex classification. These three are all very accurate, however neural networks take longer to train compared to a decision forest.

- One-versus-all multiclass algorithms can also be used. In these algorithms a two-class classifier is trained for each possible category. The results from the binary classifiers are used to classify the system as a whole.

Discovering Structure - K-Means

K-Means clustering is an algorithm which finds data in a dataset which are similar, and groups them together. In this way it helps to find structure by describing how large groups of data are and what their relations might be.

- K-Means is unsupervised, meaning it does not use pre-trained labels to learn, it finds the groups for you!
- This algorithm works by defining random centroids used to form groups, then puts data points into a group with the centroid who is the most similar. The algorithm then assesses how accurate the groups are and looks to improve until the model doesn’t change much from iteration to iteration.
- K-Means is the standard for clustering, but k-median, mean-shift, Hierarchical Clustering, and density-based spatial clustering.

Finding Unusual Data Points - Anomaly Detection

When our goal is to find unusual data points in a data set we use anomaly detection algorithms.

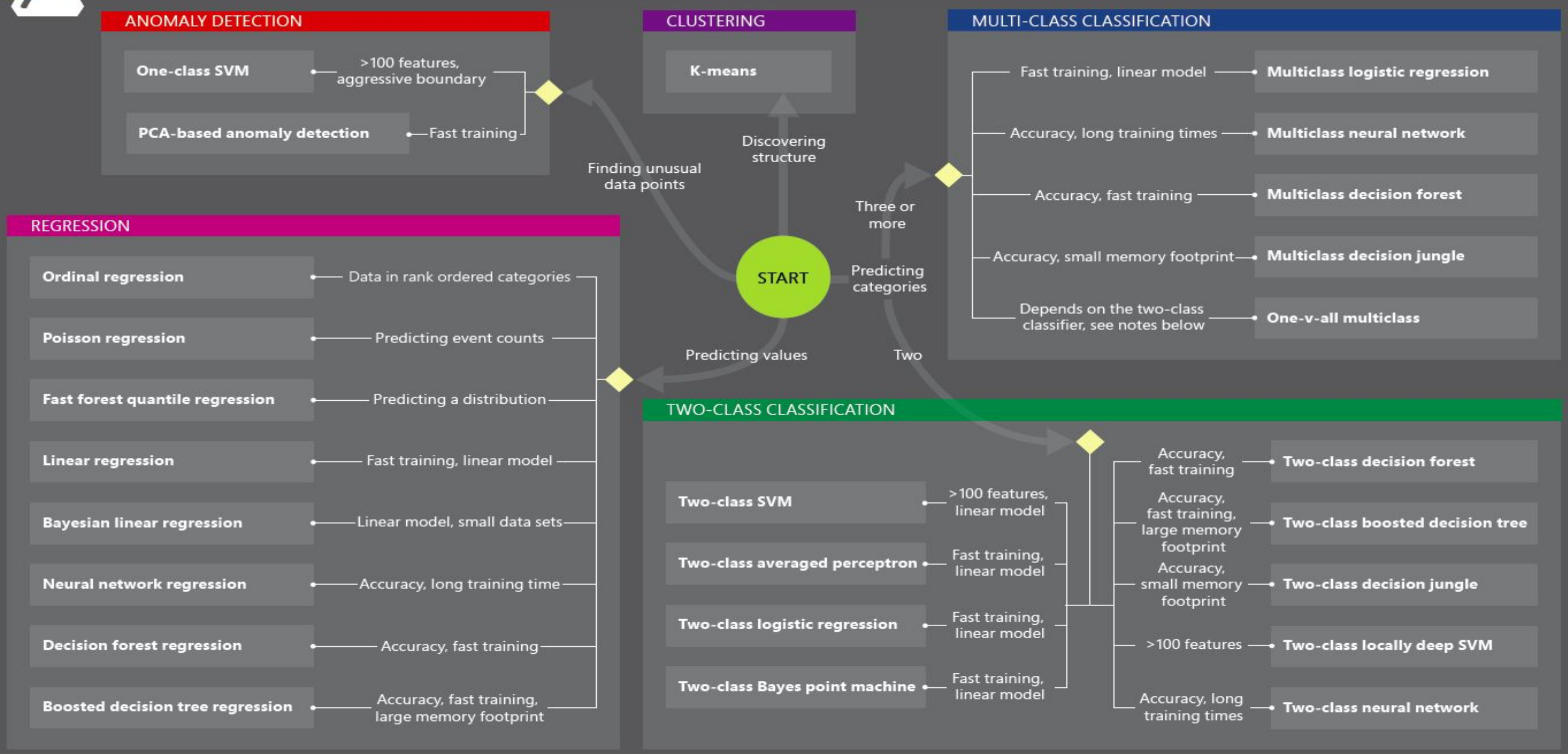
- Two useful anomaly detection methods are the one-class SVM and PCA-based anomaly detection algorithms.

- A one-class SVM algorithm is useful when you have a lot of “normal” data, but maybe not as many anomalies or you don’t know what an anomaly might look like. It is trained on data with only one “class”, the “normal” data. One-class SVM algorithms are good for data sets with a large number of features, and when you’d like an aggressive boundary.

- PCA-Based anomaly detection is better when your training data is known to be unbalanced with anomalies. It figures out which members of the data are normal and which members are anomalies. It uses dimension reduction to simplify the data, making training quicker.

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



Created by the Azure Machine Learning Team
Email: AzurePoster@microsoft.com
Download this poster: http://aka.ms/MLCheatSheet
Microsoft

Summary

- Machine Learning uses algorithms and training data to construct a model which is able to come to an insightful conclusion given a new input.
- Programs can learn through three primary ways:
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
- Common Machine Learning problems include:
  - Regression
  - Classification
  - Clustering
  - Association Rule Learning
- Dimensionality reduction finds the most meaningful data features, simplifying the data
- There are nearly endless algorithms and methods to implement machine learning, choose the right one!

Resources

https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781783558742/1/ch01lvl1sec12/taxonomy-of-machine-learning-algorithms

https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Machine Learning By Analogy by Colleen M. Farrelly

http://blog.adnanmasood.com/2016/05/26/the-elusive-search-of-approachable-taxonomy-for-machine-learning-algorithms/

https://en.wikipedia.org/wiki/Machine_learning

https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec

https://www.cs.cmu.edu/~tom/10701_sp11/slides/Kernels_SVM_04_7_2011-ann.pdf

https://developers.google.com/machine-learning/problem-framing/cases

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/pca-based-anomaly-detection

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/one-class-support-vector-machine