

Выбор и создание модели обучения

[Искусственный интеллект и экспертные системы]

Капырин Николай, старший преподаватель каф. 305
Москва, 2018

Московский Авиационный Институт

Сколько всего моделей обучения?

Сколько всего моделей обучения?

Алгоритмы обучения

Сколько всего моделей обучения?

Алгоритмы обучения

Выбор модели

Сколько всего моделей обучения?

Обучаемость (в приборных комплексах) можно понимать как способность применяемых алгоритмов обеспечивать **эмпирическое обобщение**.

Обучаемость (в приборных комплексах) можно понимать как способность применяемых алгоритмов обеспечивать **эмпирическое обобщение**.

Неформально, машинное обучение можно представить как процесс нахождения неизвестного решающего правила (или неизвестной целевой функции) по некоторой начальной информации, которая не является полной.

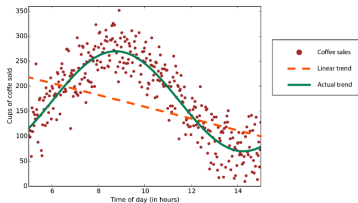
Два основных типа проблем в математике:

1. **математический анализ** – определить математические свойства (мощность, полноту, существование решений и др.) заданных математических объектов (например, множеств, семейств функций, уравнений)

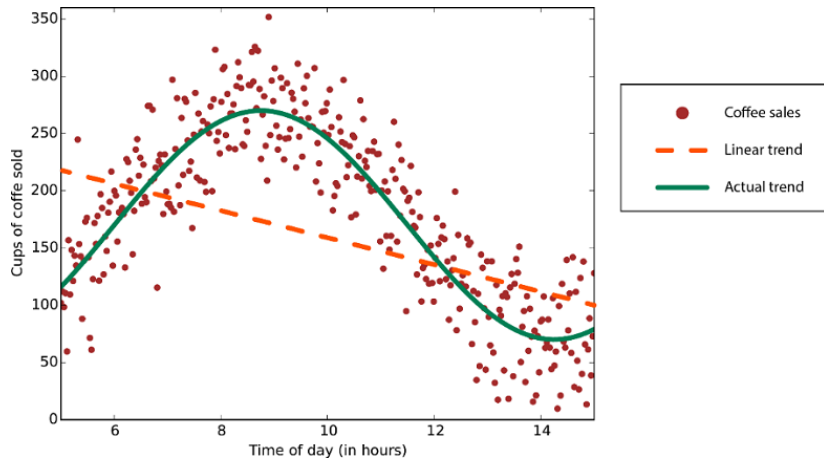
Два основных типа проблем в математике:

1. **математический анализ** – определить математические свойства (мощность, полноту, существование решений и др.) заданных математических объектов (например, множеств, семейств функций, уравнений)
2. **математический синтез** – найти математический объект, удовлетворяющий данным свойствам (в частности, нахождение неизвестного решающего/классифицирующего правила в задаче машинного обучения)

Согласованной с обучающей информацией или *корректной на ней* называется любая функция, которая в точках, входящих в эту обучающую информацию, принимает точно такие же значения, какие содержатся в примерах из этой обучающей информации.



Пример про кофе



В чем отличие обучения от нахождения произвольной корректной на данной обучающей информации функции?

В чем отличие обучения от нахождения произвольной корректной на данной обучающей информации функции?

- Число корректных на обучающей выборке рекурсивных функций сколь угодно велико.

В чем отличие обучения от нахождения произвольной корректной на данной обучающей информации функции?

- Число корректных на обучающей выборке рекурсивных функций сколь угодно велико.
- Для любой обучающей выборки ограниченной длины, существует сколь угодно много корректных на этой выборке алгоритмов обучения.

В чем отличие обучения от нахождения произвольной корректной на данной обучающей информации функции?

- Число корректных на обучающей выборке рекурсивных функций сколь угодно велико.
- Для любой обучающей выборки ограниченной длины, существует сколь угодно много корректных на этой выборке алгоритмов обучения.
- Если обучение ставит целью **приблизить порождающую модель явления**, то истинная модель должна быть единой

В чем отличие обучения от нахождения произвольной корректной на данной обучающей информации функции?

- Число корректных на обучающей выборке рекурсивных функций сколь угодно велико.
- Для любой обучающей выборки ограниченной длины, существует сколь угодно много корректных на этой выборке алгоритмов обучения.
- Если обучение ставит целью **приблизить порождающую модель явления**, то истинная модель должна быть единой, и решению сопутствует извлечение дополнительных сведений из/о выборке.

- информацию о множестве (допустимых) объектов

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации
- в каком классе решающих правил будет отыскиваться решение

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации
- в каком классе решающих правил будет отыскиваться решение
- какие дополнительные свойства множества допустимых объектов и функций должны быть учтены

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации
- в каком классе решающих правил будет отыскиваться решение
- какие дополнительные свойства множества допустимых объектов и функций должны быть учтены
- как оценивать качество обучения

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации
- в каком классе решающих правил будет отыскиваться решение
- какие дополнительные свойства множества допустимых объектов и функций должны быть учтены
- как оценивать качество обучения
- как определять, существует ли возможность достижения требуемого качества обучения при перечисленных условиях (имеет ли место обучаемость)

- информацию о множестве (допустимых) объектов
- о каком неизвестном решающем правиле или функции идёт речь
- что предоставляется в качестве начальной информации
- в каком классе решающих правил будет отыскиваться решение
- какие дополнительные свойства множества допустимых объектов и функций должны быть учтены
- как оценивать качество обучения
- как определять, существует ли возможность достижения требуемого качества обучения при перечисленных условиях (имеет ли место обучаемость)
- как оценивать число обучающих примеров, требуемых для достижения нужного качества обучения

Теория равномерной сходимости (Вепника-Червоненкиса, VC dimension, VCD) – понятие ёмкости класса решающих правил, в котором отыскивается классифицирующий алгоритм (характеристика сложности функциональных семейств).

Теория равномерной сходимости (Вапника-Червоненкиса, VC dimension, VCD) – понятие ёмкости класса решающих правил, в котором отыскивается классифицирующий алгоритм (характеристика сложности функциональных семейств).

Теория вероятно почти корректного обучения (probably approximately correct learning, PAC) – оценивается ёмкость класса решающих правил, берётся эмпирическая частота ошибок, формируется функционал и оценивается возможность найти обучаемую модель заданной точности в заданном классе.

Теория равномерной сходимости (Вапника-Червоненкиса, VC dimension, VCD) – понятие ёмкости класса решающих правил, в котором отыскивается классифицирующий алгоритм (характеристика сложности функциональных семейств).

Теория вероятно почти корректного обучения (probably approximately correct learning, PAC) – оценивается ёмкость класса решающих правил, берётся эмпирическая частота ошибок, формируется функционал и оценивается возможность найти обучаемую модель заданной точности в заданном классе.

Критерий Акаике (Akaike's information criterion, AIC) – критерий выбора из класса параметризованных регрессионных моделей с функцией штрафа за количество параметров.

- критерий средней ошибки на контрольных данных

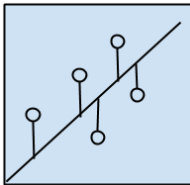
- критерий средней ошибки на контрольных данных
- критерий скользящего контроля (чтобы результат не зависел от дискретизации)

- критерий средней ошибки на контрольных данных
- критерий скользящего контроля (чтобы результат не зависел от дискретизации)
- критерии непротиворечивости/помехоустойчивость (приведение к одинаковой модели по различным подвыборкам)

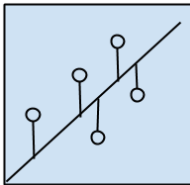
- критерий средней ошибки на контрольных данных
- критерий скользящего контроля (чтобы результат не зависел от дискретизации)
- критерии непротиворечивости/помехоустойчивость (приведение к одинаковой модели по различным подвыборкам)
- критерии регуляризации (штрафы за выход параметров алгоритма за выделенную область)

- критерий средней ошибки на контрольных данных
- критерий скользящего контроля (чтобы результат не зависел от дискретизации)
- критерии непротиворечивости/помехоустойчивость (приведение к одинаковой модели по различным подвыборкам)
- критерии регуляризации (штрафы за выход параметров алгоритма за выделенную область)
- оценка обобщающей способности (VC, AIC, PAC, BIC)

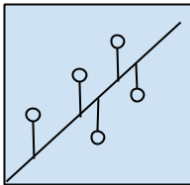
Алгоритмы обучения



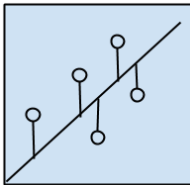
Регрессионный анализ – метод моделирования измеряемых данных и исследования их свойств.



Регрессионный анализ – метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной).



Регрессионный анализ – метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель есть функция независимой переменной и параметров с добавленной случайной переменной.



Регрессионный анализ – метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель есть функция независимой переменной и параметров с добавленной случайной переменной.

Регрессионным анализом называется поиск такой функции f , которая описывает зависимость $E(y|x) = f(x)$, где, например, $y = f(x) + v$ (v – случайный сигнал).

- Ordinary Least Squares Regression (OLSR)
- Linear Regression
- Logistic Regression
- Stepwise Regression
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)

Линейная регрессия (англ. Linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости.

Линейная регрессия (англ. Linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости.

1. Гомоскедастичность (постоянная или одинаковая дисперсия) или отсутствие гетероскедастичности случайных ошибок модели

Линейная регрессия (англ. Linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости.

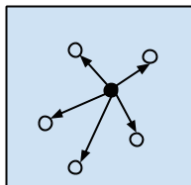
1. Гомоскедастичность (постоянная или одинаковая дисперсия) или отсутствие гетероскедастичности случайных ошибок модели
2. Отсутствие автокорреляции случайных ошибок

Линейная регрессия (англ. Linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости.

1. Гомоскедастичность (постоянная или одинаковая дисперсия) или отсутствие гетероскедастичности случайных ошибок модели
2. Отсутствие автокорреляции случайных ошибок
3. Факторы предполагаются детерминированными (нестохастическими)

Линейная регрессия (англ. Linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости.

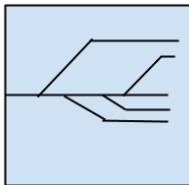
1. Гомоскедастичность (постоянная или одинаковая дисперсия) или отсутствие гетероскедастичности случайных ошибок модели
2. Отсутствие автокорреляции случайных ошибок
3. Факторы предполагаются детерминированными (нестохастическими)
4. Предполагается что отсутствует полная коллинеарность факторов



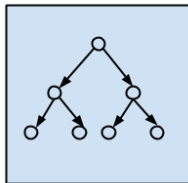
- k-Nearest Neighbour (kNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)

Метод ближайших соседей – простейший метрический классификатор, основанный на оценивании сходства объектов.

Метод ближайших соседей – простейший метрический классификатор, основанный на оценивании сходства объектов. Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки.



- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Elastic Net
- Least-Angle Regression (LARS)



- Classification and Regression Tree (CART)
- Iterative Dichotomiser 3 (ID3)
- C4.5 and C5.0 (different versions of a powerful approach)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- M5
- Conditional Decision Trees

Дерево принятия решений

(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Дерево принятия решений

(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Структура дерева представляет собой «листья» и «ветки».

Дерево принятия решений

(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Структура дерева представляет собой «**листья**» и «**ветки**». На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах – атрибуты, по которым различаются случаи.

Дерево принятия решений

(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Структура дерева представляет собой «**листья**» и «**ветки**». На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах – атрибуты, по которым различаются случаи.

Чтобы классифицировать **новый случай**, надо спуститься по дереву до листа и выдать соответствующее значение.

Дерево принятия решений

(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Структура дерева представляет собой «**листья**» и «**ветки**». На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах – атрибуты, по которым различаются случаи.

Чтобы классифицировать **новый случай**, надо спуститься по дереву до листа и выдать соответствующее значение.

Широко используются в интеллектуальном анализе данных.

Дерево принятия решений

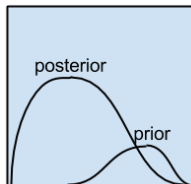
(дерево классификации, регрессионное дерево)

Средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей.

Структура дерева представляет собой «**листья**» и «**ветки**». На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах – атрибуты, по которым различаются случаи.

Чтобы классифицировать **новый случай**, надо спуститься по дереву до листа и выдать соответствующее значение.

Широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе.



- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Belief Network (BBN)
- Bayesian Network (BN)

Байесовский классификатор – простой вероятностный классификатор, основанный на применении Теоремы Байеса со строгими (наивными) предположениями о независимости.

Байесовский классификатор – простой вероятностный классификатор, основанный на применении Теоремы Байеса со строгими (наивными) предположениями о независимости.

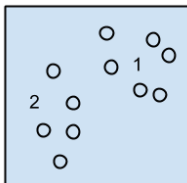
- Могут обучаться очень эффективно

Байесовский классификатор – простой вероятностный классификатор, основанный на применении Теоремы Байеса со строгими (наивными) предположениями о независимости.

- Могут обучаться очень эффективно
- Несмотря на наивный вид и очень упрощённые условия, БК часто отлично работают во многих сложных ситуациях.

Байесовский классификатор – простой вероятностный классификатор, основанный на применении Теоремы Байеса со строгими (наивными) предположениями о независимости.

- Могут обучаться очень эффективно
- Несмотря на наивный вид и очень упрощённые условия, БК часто отлично работают во многих сложных ситуациях.
- малое количество данных для обучения, необходимых для оценки параметров, требуемых для классификации.



- k-Means
- k-Medians
- Expectation Maximisation (EM)
- Hierarchical Clustering

Метод опорных векторов (SVM, support vector machine)

Набор алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа.

Метод опорных векторов (SVM, support vector machine)

Набор алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Принадлежит к семейству линейных классификаторов, может также рассматриваться как специальный случай регуляризации по Тихонову.

Метод опорных векторов (SVM, support vector machine)

Набор алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Принадлежит к семейству линейных классификаторов, может также рассматриваться как специальный случай регуляризации по Тихонову.

Основная идея – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве.

Метод опорных векторов (SVM, support vector machine)

Набор алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Принадлежит к семейству линейных классификаторов, может также рассматриваться как специальный случай регуляризации по Тихонову.

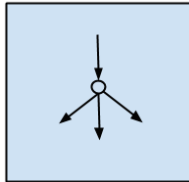
Основная идея – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве.

$(A,B) \longrightarrow C$

$(D,E) \longrightarrow F$

$(A,E) \longrightarrow G$

- Apriori algorithm
- Eclat algorithm



- Perceptron
- Back-Propagation
- Hopfield Network
- Radial Basis Function Network (RBFN)

Искусственная нейронная сеть (artificial neural network, ANN, нейронная сеть) – это математическая модель, а также её программные или аппаратные реализации, построенная в некотором смысле по образу и подобию сетей нервных клеток живого организма.

Искусственная нейронная сеть (artificial neural network, ANN, нейронная сеть) – это математическая модель, а также её программные или аппаратные реализации, построенная в некотором смысле по образу и подобию сетей нервных клеток живого организма.

ИНС представляют собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов).

Искусственная нейронная сеть (artificial neural network, ANN, нейронная сеть) – это математическая модель, а также её программные или аппаратные реализации, построенная в некотором смысле по образу и подобию сетей нервных клеток живого организма.

ИНС представляют собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов).

Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах).

Искусственная нейронная сеть (artificial neural network, ANN, нейронная сеть) – это математическая модель, а также её программные или аппаратные реализации, построенная в некотором смысле по образу и подобию сетей нервных клеток живого организма.

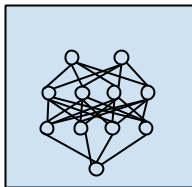
ИНС представляют собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов).

Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам.

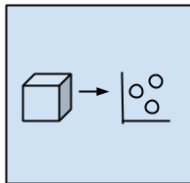
Искусственная нейронная сеть (artificial neural network, ANN, нейронная сеть) – это математическая модель, а также её программные или аппаратные реализации, построенная в некотором смысле по образу и подобию сетей нервных клеток живого организма.

ИНС представляют собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов).

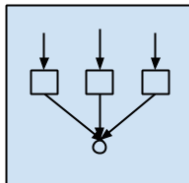
Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И, тем не менее, будучи соединёнными в достаточно большую сеть с управляемым взаимодействием, такие локально простые процессоры вместе способны выполнять довольно сложные задачи.



- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Convolutional Neural Network (CNN)
- Stacked Auto-Encoders



- Principal Component Analysis/Regression (PCA, PCR)
- Partial Least Squares Regression (PLSR)
- Sammon Mapping
- Multidimensional Scaling (MDS)
- Projection Pursuit
- Linear/Mixture Discriminant Analysis (LDA, MDA)
- Quadratic/Flexible Discriminant Analysis (QDA, FDA)



- Boosting
- Bootstrapped Aggregation (Bagging)
- AdaBoost
- Stacked Generalization (blending)
- Gradient Boosting Machines (GBM)
- Gradient Boosted Regression Trees (GBRT)
- **Random Forest**



Выбор модели

1. Machine_learning [Электронный ресурс] // Википедия : свободная энцикл. – Электрон. дан. – [Б. м.], 2015. – URL: http://en.wikipedia.org/wiki/Machine_learning (дата обращения: 11.11.2015).

3 Theory

4 Approaches

4.1 Decision tree learning

4.2 Association rule learning

4.3 Artificial neural networks

4.4 Inductive logic programming

4.5 Support vector machines

4.6 Clustering

4.7 Bayesian networks

4.8 Reinforcement learning

4.9 Representation learning

4.10 Similarity and metric learning

4.11 Sparse dictionary learning

4.12 Genetic algorithms

5 Applications

6 Software

Problems

Classification • Clustering • Regression •
Anomaly detection • Association rules •
Reinforcement learning •
Structured prediction •
Feature engineering • Feature learning •
Online learning •
Semi-supervised learning •
Unsupervised learning • Learning to rank
• Grammar induction

Supervised learning (classification • regression)

Decision trees • Ensembles (Bagging,
Boosting, Random forest) • k -NN •
Linear regression • Naive Bayes •
Neural networks • Logistic regression •
Perceptron •
Relevance vector machine (RVM) •
Support vector machine (SVM)

Clustering

BIRCH • Hierarchical • k -means •
Expectation-maximization (EM) •
DBSCAN • OPTICS • Mean-shift

Dimensionality reduction

Factor analysis • CCA • ICA • LDA • NMF •
PCA • t-SNE

1. Machine_learning [Электронный ресурс] // Википедия : свободная энцикл. – Электрон. дан. – [Б. м.], 2015. – URL: http://en.wikipedia.org/wiki/Machine_learning (дата обращения: 11.11.2015).

3 Theory

4 Approaches

4.1 Decision tree learning

4.2 Association rule learning

4.3 Artificial neural networks

4.4 Inductive logic programming

4.5 Support vector machines

4.6 Clustering

4.7 Bayesian networks

4.8 Reinforcement learning

4.9 Representation learning

4.10 Similarity and metric learning

4.11 Sparse dictionary learning

4.12 Genetic algorithms

5 Applications

6 Software

Problems

Classification • Clustering • Regression •
Anomaly detection • Association rules •
Reinforcement learning •
Structured prediction •
Feature engineering • Feature learning •
Online learning •
Semi-supervised learning •
Unsupervised learning • Learning to rank •
Grammar induction

Supervised learning (classification • regression)

Decision trees • Ensembles (Bagging,
Boosting, Random forest) • k -NN •
Linear regression • Naive Bayes •
Neural networks • Logistic regression •
Perceptron •
Relevance vector machine (RVM) •
Support vector machine (SVM)

Clustering

BIRCH • Hierarchical • k -means •
Expectation-maximization (EM) •
DBSCAN • OPTICS • Mean-shift

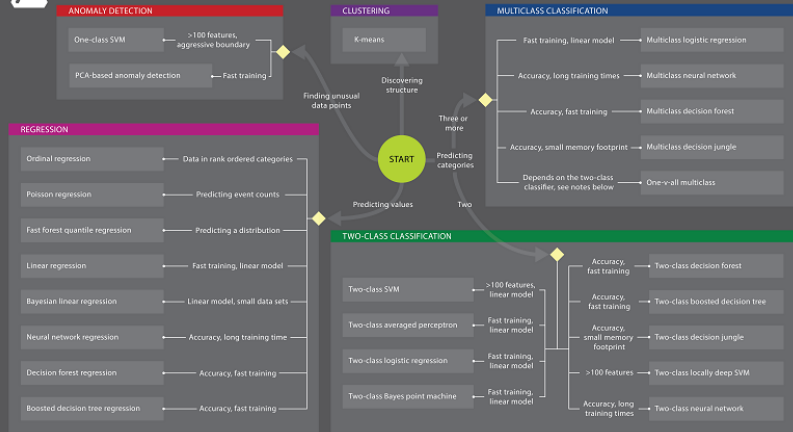
Dimensionality reduction

Factor analysis • CCA • ICA • LDA • NMF •
PCA • t-SNE



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.



Спасибо за внимание!