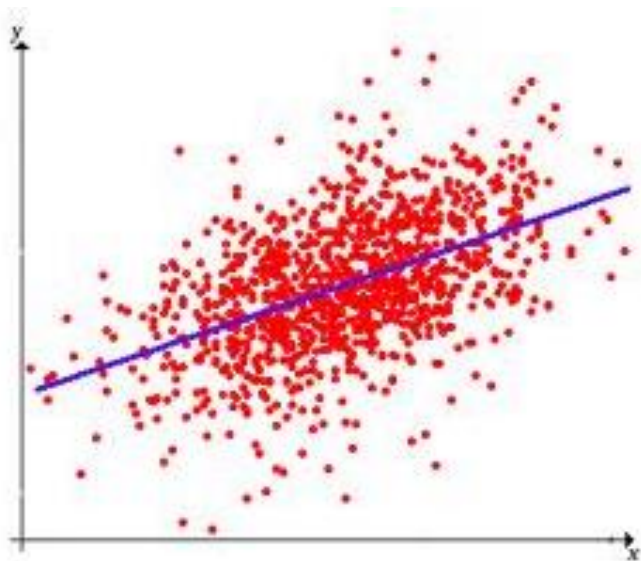


回归算法



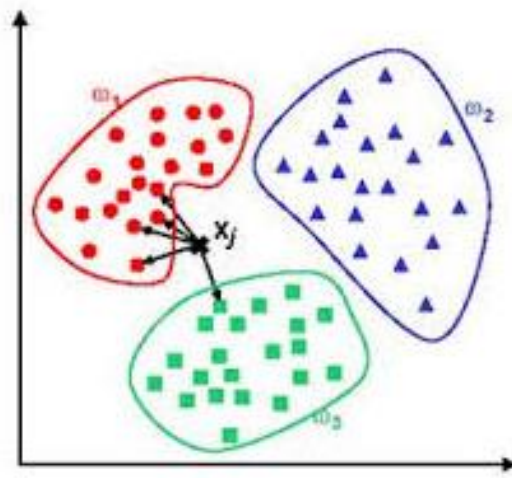
- 回归算法是试图采用对误差的衡量来探索变量之间的关系的一类算法。在机器学习领域，人们说起回归，有时候是指一类问题，有时候是指一类算法，这一点常常会使初学者有所困惑。
- 常见的回归算法包括：最小二乘法（Ordinary Least Square），逻辑回归（Logistic Regression），逐步式回归（Stepwise Regression）等。



基于实例的算法



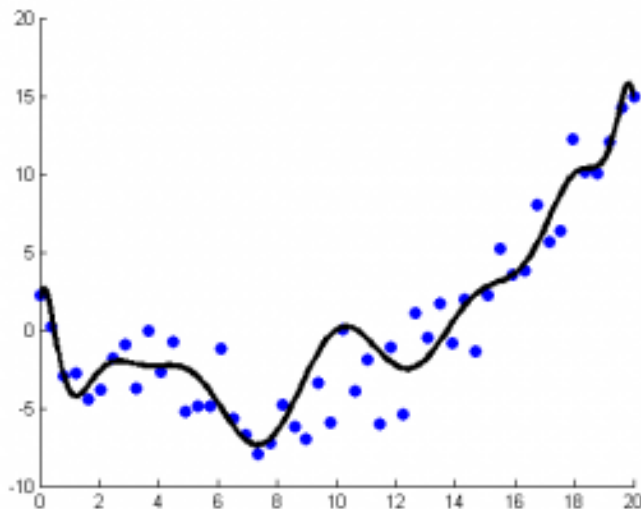
- 基于实例的算法常常用来对决策问题建立模型，这样的模型常常先选取一批样本数据，然后根据某些近似性把新数据与样本数据进行比较。通过这种方式来寻找最佳的匹配。因此，基于实例的算法常常也被称为“赢家通吃”学习或者“基于记忆的学习”。
- 常见的算法包括 k-Nearest Neighbor(KNN), 学习矢量量化 (Learning Vector Quantization, LVQ) , 以及自组织映射算法 (Self-Organizing Map , SOM)



正则化方法



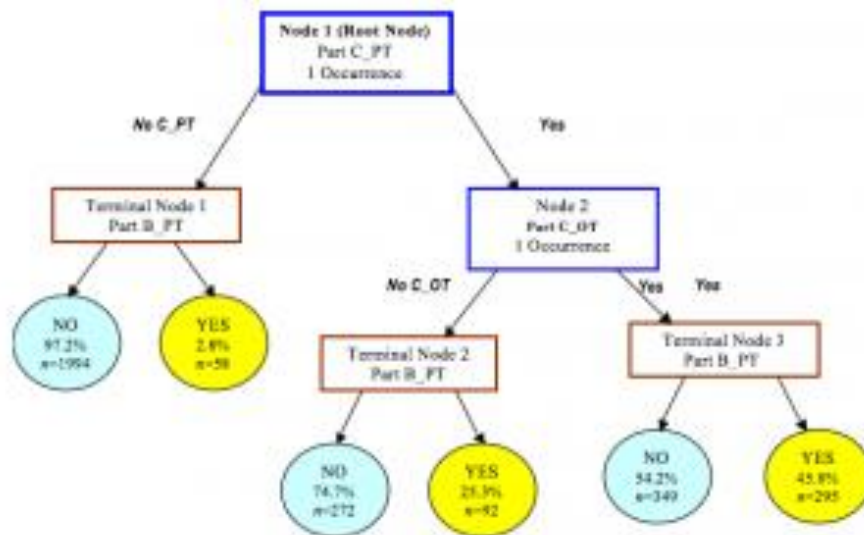
- 正则化方法是其他算法（通常是回归算法）的延伸，根据算法的复杂度对算法进行调整。正则化方法通常对简单模型予以奖励而对复杂算法予以惩罚。
- 常见的算法包括：Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO)，以及弹性网络 (Elastic Net)。



决策树学习



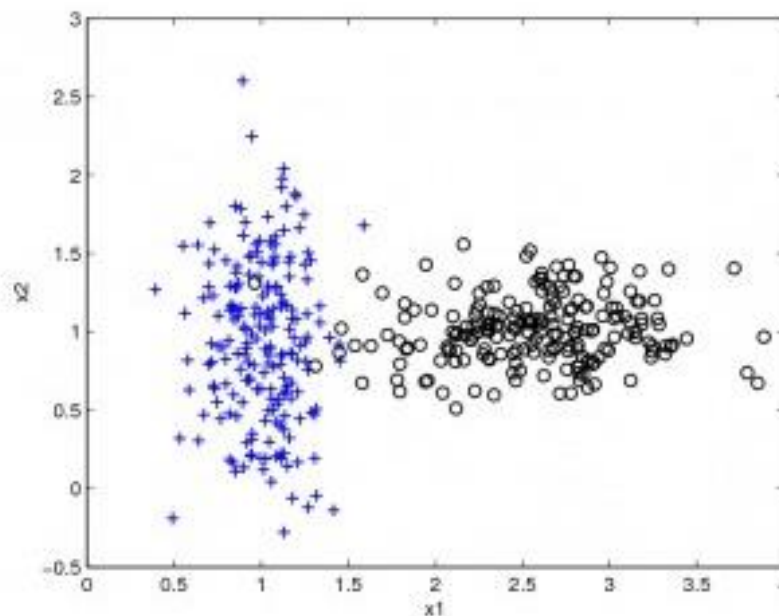
- 决策树算法根据数据的属性采用树状结构建立决策模型，决策树模型常常用来解决分类和回归问题。
- 常见的算法包括：CART, ID3, C4.5, CHAID, Decision Stump, 随机森林 (Random Forest), 多元自适应回归样条 (MARS) 以及梯度推进机 (Gradient Boosting Machine, GBM) 。



贝叶斯方法



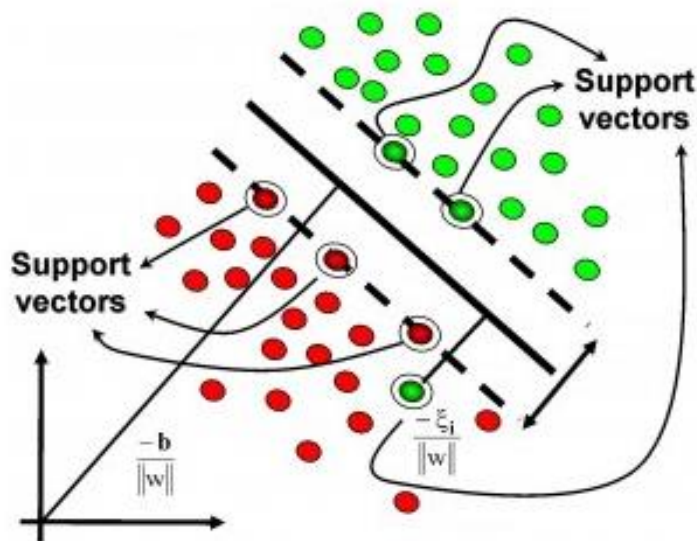
- 贝叶斯方法算法是基于贝叶斯定理的一类算法，主要用来解决分类和回归问题。
- 常见算法包括：朴素贝叶斯算法，平均单依赖估计（Averaged One-Dependence Estimators, AODE），以及Bayesian Belief Network (BBN)。



基于核的算法



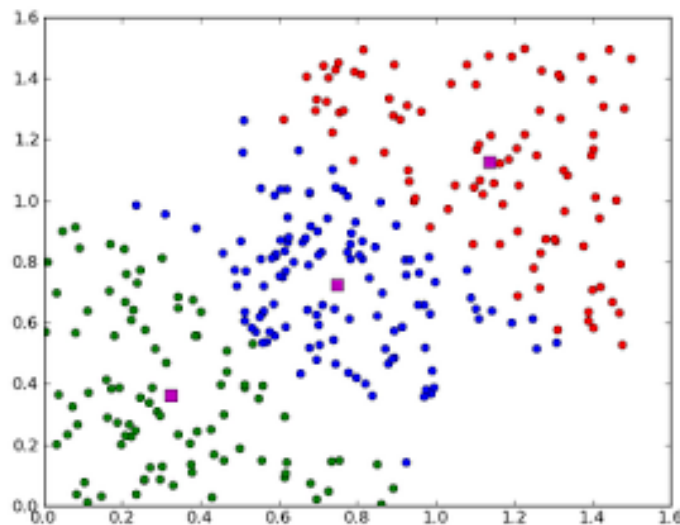
- 基于核的算法把输入数据映射到一个高阶的向量空间，在这些高阶向量空间里，有些分类或者回归问题能够更容易的解决。
- 常见的基于核的算法包括：支持向量机（Support Vector Machine, SVM），径向基函数（Radial Basis Function, RBF），以及线性判别分析（Linear Discriminate Analysis, LDA）等。



聚类算法



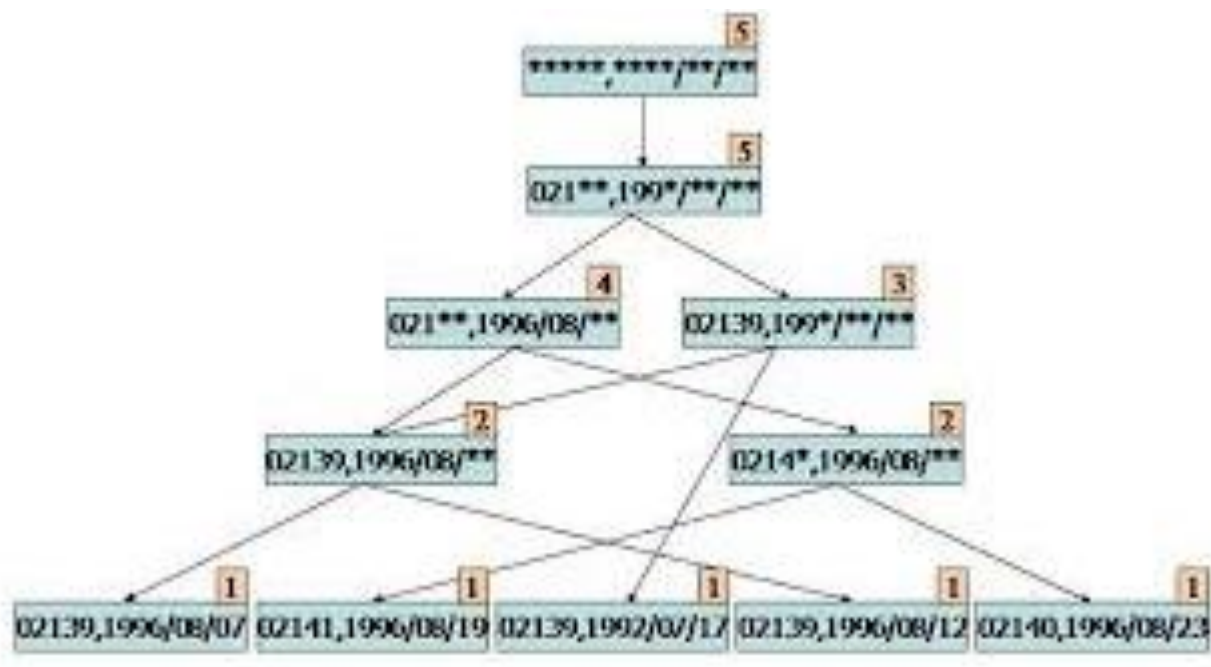
- 聚类，就像回归一样，有时候人们描述的是一类问题，有时候描述的是一类算法。聚类算法通常按照中心点或者分层的方式对输入数据进行归并。所以的聚类算法都试图找到数据的内在结构，以便按照最大的共同点将数据进行归类。
- 常见的聚类算法包括 k-Means算法以及期望最大化算法（Expectation Maximization, EM）。



关联规则学习



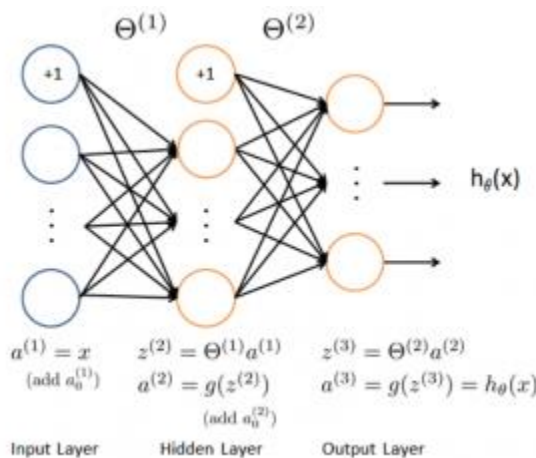
- 关联规则学习通过寻找最能够解释数据变量之间关系的规则，来找出大量多元数据集中有用的关联规则。
- 常见算法包括 Apriori算法和Eclat算法等。



人工神经网络



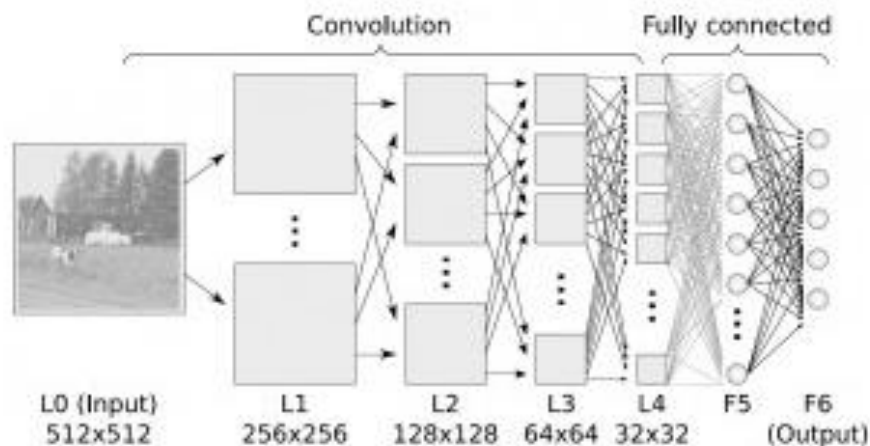
- 人工神经网络算法模拟生物神经网络，是一类模式匹配算法。通常用于解决分类和回归问题。人工神经网络是机器学习的一个庞大的分支，有几百种不同的算法。
- 重要的人工神经网络算法包括：感知器神经网络（Perceptron Neural Network），反向传递（Back Propagation），Hopfield网络，自组织映射（Self-Organizing Map, SOM）。



深度学习



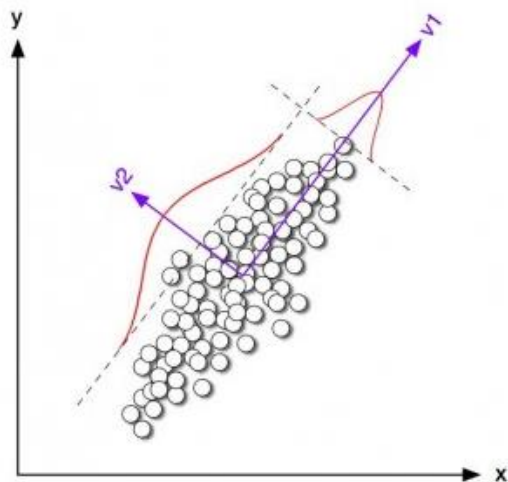
- 深度学习算法是对人工神经网络的发展。随着计算能力的提高，深度学习试图建立大得多也复杂得多的神经网络。很多深度学习的算法是半监督式学习算法，用来处理存在少量未标识数据的大数据集。
- 深度学习算法包括：受限波尔兹曼机（Restricted Boltzmann Machine, RBN），Deep Belief Networks (DBN)，卷积网络（Convolutional Network），堆栈式自动编码器（Stacked Auto-encoders）。



降低维度算法



- 像聚类算法一样，降低维度算法试图分析数据的内在结构，不过降低维度算法是以非监督学习的方式试图利用较少的信息来归纳或者解释数据。这类算法可以用于高维数据的可视化或者用来简化数据以便监督式学习使用。
- 常见算法包括：主成份分析（PCA），偏最小二乘回归（PLS），Sammon映射，多维尺度（MDS），投影追踪（Projection Pursuit）等。



集成算法



- 集成算法用一些相对较弱的学习模型独立地就同样的样本进行训练，然后把结果整合起来进行整体预测。集成算法的主要难点在于究竟集成哪些独立的较弱的学习模型以及如何把学习结果整合起来。这是一类非常强大的算法，同时也非常流行。
- 常见的算法包括：Boosting， Bootstrapped Aggregation (Bagging) ， AdaBoost， 堆叠泛化 (Stacked Generalization, Blending) ， 梯度推进机 (Gradient Boosting Machine, GBM) ， 随机森林 (Random Forest) 。

