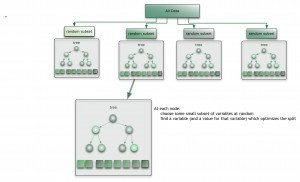
机器学习：

### **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）。

GDBT：

　GBDT(Gradient Boosting Decision Tree) 又叫 MART（Multiple Additive Regression Tree)，

### 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。常见的算法包括：主成份分析（Principle Component Analysis， PCA），偏最小二乘回归（Partial Least Square Regression，PLS）， Sammon映射，多维尺度（Multi-Dimensional Scaling, MDS）,  投影追踪（Projection Pursuit）等。K 近邻学习

<https://www.cnblogs.com/llhthinker/p/5522054.html>

面试题

https://zhuanlan.zhihu.com/p/29936999

深度学习：

深度学习算法是对人工神经网络的发展。 在近期赢得了很多关注， 特别是[百度也开始发力深度学习后](http://www.ctocio.com/ccnews/15615.html)， 更是在国内引起了很多关注。   在计算能力变得日益廉价的今天，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。常见的深度学习算法包括：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）。

RBM：

能量模型是个什么样的东西呢？直观上的理解就是，把一个表面粗糙又不太圆的小球，放到一个表面也比较粗糙的碗里，就随便往里面一扔，看看小球停在碗的哪个地方。一般来说停在碗底的可能性比较大，停在靠近碗底的其他地方也可能，甚至运气好还会停在碗口附近（这个碗是比较浅的一个碗）；能量模型把小球停在哪个地方定义为一种状态，每种状态都对应着一个能量，这个能量由能量函数来定义，小球处在某种状态的概率（如停在碗底的概率跟停在碗口的概率当然不一样）可以通过这种状态下小球具有的能量来定义（换个说法，如小球停在了碗口附近，这是一种状态，这个状态对应着一个能量E，而发生“小球停在碗口附近”这种状态的概率p，可以用E来表示，表示成p=f(E)，其中f是能量函数），这就是我认为的能量模型。

第二，能量函数。随机神经网络是根植于统计力学的。受统计力学中能量泛函的启发，引入了能量函数。能量函数是描述整个系统状态的一种测度。系统越有序或者概率分布越集中，系统的能量越小。反之，系统越无序或者概率分布越趋于均匀分布，则系统的能量越大。能量函数的最小值，对应于系统的最稳定状态。