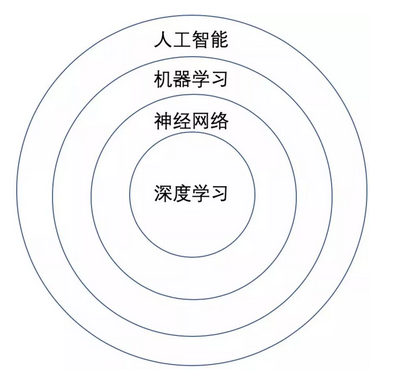
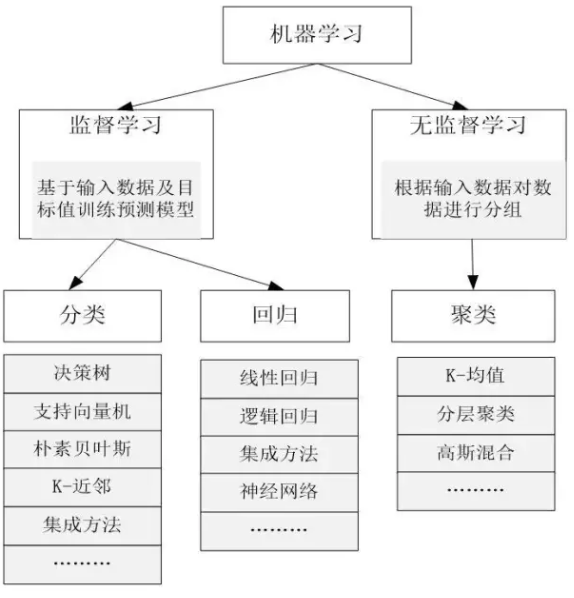
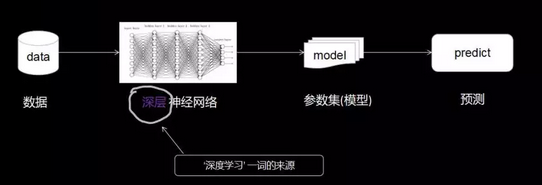
# 概述

 人工智能，机器学习和深度学习的关系：

人工智能：能够独立推理思考、有自我意识、有自己的世界观的机器（程序）。

机器学习：对于某类任务T和性能度量P，如果计算机程序在T上以P衡量的性能随着经验E而自我完善，那么就成这个计算机程序从经验E学习（机器学习）。

 机器学习按照方法主要可以分为两大类：监督学习和无监督学习（还有强化学习）。其中监督学习主要由分类和回归等问题组成，无监督学习主要由聚类和关联分析等问题组成。**深度学习则属于监督学习当中的一种**。

 深度学习：对于某类任务T和性能度量P，如果计算机程序在T上以P度量的性能随着经验E而自我完善，那么就称这个计算机程序从经验E学习。

# 算法分类

人工智能算法大致来说分为两类：基于统计的机器学习算法（Machine Learning）和深度学习算法（Deep Learning）。

总的来说，在sklearn中机器学习算法大致分类如下：

## 纯算法类

### 回归算法

回归分析是研究自变量和因变量之间关系的一种预测模型技术。这些技术应用于预测，时间序列模型和找到变量之间关系。

使用回归分析的好处：它指示出自变量和因变量之间的显著关系；它指示出多个自变量对因变量的影响。回归分析允许我们比较不同尺度的变量。

### 分类算法

### 聚类算法

### 降维算法

### 概率图模型算法

### 文本挖掘算法

### 优化算法

### 深度学习算法

## 建模类

### 模型优化

特征选取

梯度下降

交叉验证

参数调优

模型评估：准确率、召回率、F1、AUC、ROC、损失函数

### 数据预处理

标准化

异常值处理

二值化

缺失值填充：支持均值、中位数、特定值补差、多重插值

# 详细算法

## 分类算法

### LR

Logistic Regression，逻辑回归又叫逻辑分类

### SVM

Support Vector Machine，支持向量机

### NB

Naïve Bayes，朴素贝叶斯

### DT

Decision Tree，决策树

#### C4.5

#### D3

#### CART

### 集成算法

#### Bagging

#### Randon Forest

随机森林

#### GB

Gradient boosting，梯度提升

#### GBDT

Gradient Boosting Decision Tree

#### AdaBoost

#### Xgboost

### 最大熵模型

## 回归算法

### LR

Liniear Regression，线性回归

### SVR

支持向量机回归

### RR

Ridge Regression，岭回归

## 聚类算法

### KNN

### Kmeans算法

### 层次据诶

### 密度聚类

## 降维算法

### SGD

随机梯度下降

## 概率图模型算法

### 贝叶斯网络

### HMM

### CRF

条件随机场

## 优化算法

### 正则化

#### L1正则化

#### L2正则化

## 深度学习算法

### BP

### CNN

### DNN

### RNN

### GAN

### LSTM

# 应用

## 文本挖掘算法

### 模型

#### LDA

Latent Dirichlet Allocation，主题生成模型

#### 最大熵模型

### 关键词提取

#### tf-idf

#### bm25

#### textrank

#### pagerank

#### 左右熵

#### 互信息

### 词法分析

#### 分词

##### HMM

隐马尔可夫

##### CRF

条件随机场

#### 词性标注

#### 命名实体识别

### 句法分析

#### 句法结构分析

#### 依存句法分析

### 文本向量化

#### tf-idf

#### word2vec

#### doc2vec

#### cw2vec

### 距离计算

#### 欧氏距离

#### 相似度计算

## 语音识别算法

## 图像识别算法