**人工智能知识点**

一、基本概念和知识

1、监督学习：线性回归、对数几率回归、决策树

2、非监督学习：

3、强化学习：

4、迁移学习：

**5、特征工程：**

6、github账户：

7、csdn账户：

二、传统机器学习算法

1、**LR（逻辑回归）：**

2、**决策树：**

3、**决策森林：**

4、**GBDT：**

5、**信息熵、信息增益、信息增益率：**

6、线性回归、最小二乘法：

7、L1、L2正则化：

8、朴素贝叶斯分类器：

9、SVM支持向量机、核函数：

10、聚类算法k-means、均值漂移、图团体检测(Graph Community Detection)

11、xgboost思想：

12、faiss算法：

**13、itemCF、userCF：**

14、PageRank：

15、正则化、归一化、标准化：

 归一化（Normalization）

    （1）把数据变为（0，1）之间的小数。主要是为了方便数据处理，因为将数据映射到0～1范围之内，可以使处理过程更加便捷、快速。

    （2）把有量纲表达式变换为无量纲表达式，成为纯量。经过归一化处理的数据，处于同一数量级，可以消除指标之间的量纲和量纲单位的影响，提高不同数据指标之间的可比性。

     主要算法：

    （1）线性转换，即min-max归一化（常用方法）： y=(x-min)/(max-min)

    （2）对数函数转换：y=log10(x)

（3）反余切函数转换：y=atan(x)\*2/PI

标准化（Standardization）

    数据的标准化是将数据按比例缩放，使之落入一个小的特定区间。

    主要方法：

    （1）z-score标准化，即零-均值标准化（常用方法）

    y=(x-μ)/σ，是一种统计的处理，基于正态分布的假设，将数据变换为均值为0、标准差为1的标准正态分布。但即使数据不服从正态分布，也可以用此法。特别适用于数据的最大值和最小值未知，或存在孤立点。

    （2）小数定标标准化

     y=x/10^j  （j确保max(|y|)<1）通过移动x的小数位置进行标准化

    （3）对数Logistic模式： y=1/(1+e^(-x))

正则化（Regularization）

     用一组与原不适定问题相“邻近”的适定问题的解，去逼近原问题的解，这种方法称为正则化方法。如何建立有效的正则化方法是反问题领域中不适定问题研究的重要内容。通常的正则化方法有基于变分原理的Tikhonov 正则化、各种迭代方法以及其它的一些改进方法。

     总的来说，归一化是为了消除不同数据之间的量纲，方便数据比较和共同处理，比如在神经网络中，归一化可以加快训练网络的收敛性；标准化是为了方便数据的下一步处理，而进行的数据缩放等变换，并不是为了方便与其他数据一同处理或比较，比如数据经过零-均值标准化后，更利于使用标准正态分布的性质，进行处理；正则化而是利用先验知识，在处理过程中引入正则化因子(regulator)，增加引导约束的作用，比如在逻辑回归中使用正则化，可有效降低过拟合的现象。

三、神经网络

1、神经网络

2、BP算法

3、激活函数

4、CNN

5、Dropout

**6、Optimizer优化器：**

（1）梯度下降Gradient Descent（GD）

（2）批量梯度下降Batch Gradient Descent（BGD）

（3）随机梯度下降Stochastic Gradient Descent（SGD）

（4）自适应学习率优化算法：AdaGrad算法，RMSProp算法，Adam算法以及AdaDelta算法

（5）动量优化Momentum

（6）牛顿加速梯度（NAG, Nesterov accelerated gradient）算法

四、编程语言

1、python

2、c++

3、java

五、平台框架

1、numpy

2、pandas

3、spark

4、redis

5、linux

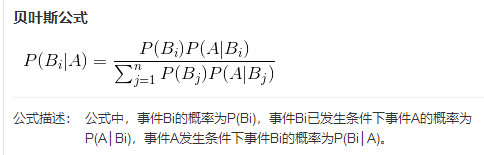
6、sql

7、**tensorflow：**

8、sklearn

六、数学知识

1、**概率论**



2、线性代数