1. **L1正则化和L2正则化有什么区别？**L1正则化和L2正则化的先验分布

来源：vivo 算法岗： <https://www.nowcoder.com/discuss/198053>

来源 - 度小满 数据分析：<https://www.nowcoder.com/discuss/189529>

课程相关内容：监督学习入门I - 正则化

回答：正则化的目的是防止模型过拟合，具体做法是在损失函数后面加正则项。其中最常用的两种正则化方法是L1正则化和L2正则化，L1是模型各个参数的绝对值之和，L2是模型各个参数的平方和的开方值。L1会趋向于产生少量的特征，而其他的特征都是0。L2会选择更多的特征，这些特征都会接近于0。   L1的先验分布是拉普拉斯分布，L2的先验分布是高斯分布。

1. **kmeans的详细过程，以及如何确定K值？**

来源 - 京东 数据分析：<https://www.nowcoder.com/discuss/182126>

课程相关内容：P1 机器学习入门 - K-means

回答：对于K-means中K的选择，通常有四种方法。

* 按需选择
* 观察法
* 手肘法

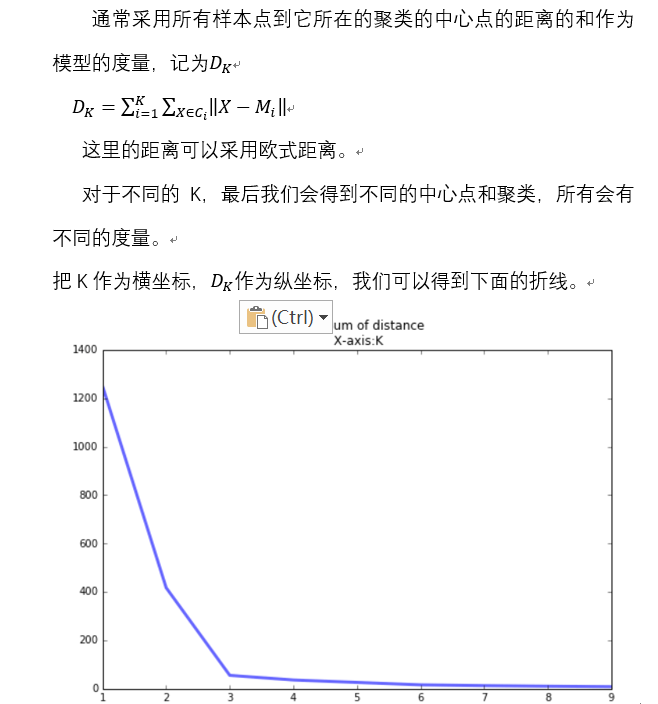
**一、按需选择**

简单地说就是按照建模的需求和目的来选择聚类的个数。比如说，一个游戏公司想把所有玩家做聚类分析，分成顶级、高级、中级、菜鸟四类，那么K=4

**二、观察法**

通过降维算法把数据降到二维或者三维，然后观察数据的散点图。

**三、手肘法**

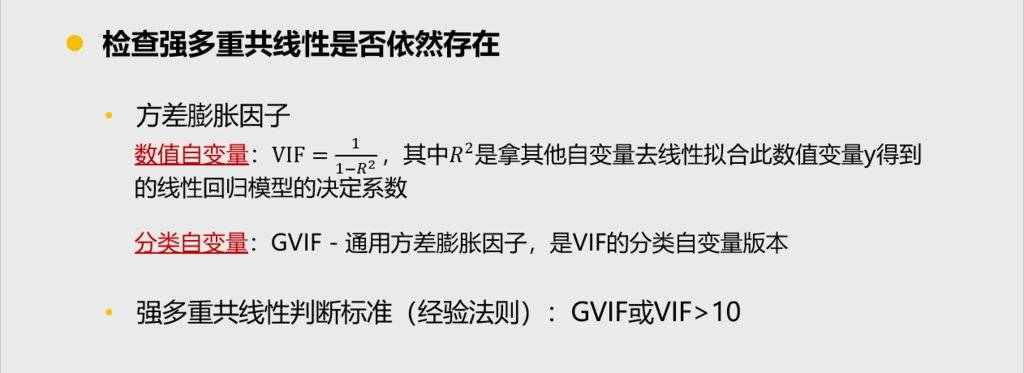


很显然K越大，距离和越小。但是我们注意到K=3是一个拐点，就像是我们的肘部一样，K=1到3下降很快，K=3之后趋于平稳。**手肘法认为这个拐点就是最佳的K。**

1. **如何排除多重共线性？**

来源 - 京东 数据分析：<https://www.nowcoder.com/discuss/182126>

课程相关内容：P2 监督学习入门 - 线性回归，逻辑回归

回答： 

1. **AUC跟准确度、F1等指标相比有什么好处？**

来源 - 追一科技 算法岗：<https://www.nowcoder.com/discuss/217585>

课程相关内容：P2 监督学习入门II-模型效果评估

回答：

F1为Precision和Recall的调和平均 ，数值上一般接近于二者中的较小值。

AUC：任意取一个正样本和负样本，正样本得分大于负样本的概率。

两种计算AUC的方法

1：绘制ROC曲线，ROC曲线下面的面积就是AUC的值

2：假设总共有（m+n）个样本，其中正样本m个，负样本n个，总共有m\*n个样本对，计数，正样本预测为正样本的概率值大于负样本预测为正样本的概率值记为1，累加计数，然后除以（m\*n）就是AUC的值

1. **缺失值的处理方法**

来源-优酷 算法岗：<https://www.nowcoder.com/discuss/215907>

课程相关内容：P2 毕业实战设计

回答：

1.删除

1.1 删除缺失特征较多的样本  
1.2 删除缺失比例较大的特征

2.填补

2.1 特殊值填充。将空值作为一种特殊的属性值来处理，它不同于其他的任何属性值。

2.2 均值填充。当特征为数值型变量时，用该特征在其他所有样本的取值的平均值来填充该缺失的属性值；当特征为类别型变量时，用该特征在其他所有样本的取值的众数来填充该缺失的属性值。

2.3 条件均值填充。与均值填充不同的是，取同类（相同标签）样本的平均值或众数来填充。

2.4 用算法拟合进行填充。将缺失特征作为因变量，无缺失特征作为自变量，训练模型对缺失特征进行预测，用预测值进行填充。

**5.讲一讲SVM (建模思想，误差函数推导，核，优化), 合页损失函数**

来源：京东[算法岗：https://www.nowcoder.com/discuss/19](http://算法岗：https:/www.nowcoder.com/discuss/199755)9755

阿里妈妈 算法岗：<https://www.nowcoder.com/discuss/172988>

课程相关内容：监督学习入门-SVM

回答：...

1. **SVM和lr的区别和各自优缺点**

来源 - 字节跳动 算法岗：<https://www.nowcoder.com/discuss/204240>

课程相关内容：监督学习入门-SVM,LR

回答：

一句话概括LR：逻辑回归假设数据服从伯努利分布,通过极大化似然函数的方法，运用梯度下降来求解参数，来达到将数据二分类的目的。

一句话概括SVM：找到一个超平面，使得两类数据点到该超平面的最近距离最大化，这些离超平面最近的点就被称为支持向量。

联系：1.不考虑核函数，两者都是线性分类算法

2.都是判别式模型

区别： 1.目标函数不同，代表了不同的假设前提，也就代表了不同的分类原理。LR基于概率，SVM基于距离。

2.SVM 只考虑局部的边界线附近的点，LR 考虑全局，远离的点对边界线的确定也起作用。

3.线性 SVM 依赖数据表达的距离测度，所以训练前需要先对数据做归一化, LR 则不受影响。

LR优点：

1.形式简单，模型的可解释性非常好。

2.训练速度较快。

LR缺点：

1.较难处理线性不可分数据。

SVM优点：

1.有大量的核函数可以使用，从而可以很灵活的来解决各种非线性的分类问题。

2.仅仅使用一部分支持向量来做超平面的决策，无需依赖全部数据。

SVM缺点：

1.SVM在样本量非常大，核函数映射维度非常高时，计算量过大，不太适合使用。

1. **LR为什么要用sigmoid？**

来源 - 字节跳动 算法岗：<https://www.nowcoder.com/discuss/204240>

课程相关内容：监督学习入门-LR

回答：

1.Sigmoid能够将输出映射到[0,1]之间。正好这个是概率的范围。

2. Sigmoid是连续光滑的。

3.广义线性模型理论和伯努利分布推导出的函数形式和sigmoid函数相同。

7.1**LR为什么要离散特征？**

来源-平安银行 算法岗：<https://www.nowcoder.com/discuss/163388>

课程相关内容：监督学习入门-LR

回答：

1.离散化后的特征对异常数据有很强的鲁棒性

2. 稀疏向量内积乘法运算速度快

3.离散化后可以进行特征交叉，由M+N个变量变为M\*N个变量，进一步引入非线性，提升表达能力。

**8.分类问题的loss为什么选交叉熵，MSE可以吗？**

[美团-算法岗：https://www.nowcoder.com/discuss/186231](https://www.nowcoder.com/discuss/186231)

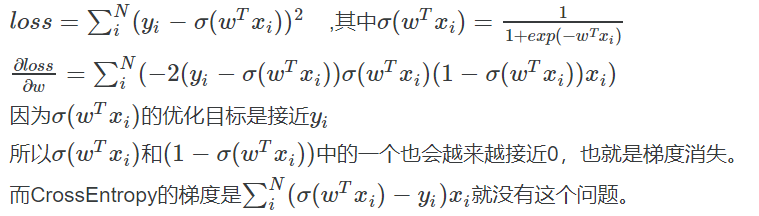
课程相关内容：监督学习入门-逻辑回归

回答：

1.当损失函数是MSE的时候，不容易找到最小值点

2.用MSE作为逻辑回归的损失函数，那么这个损失函数本身不是凸函数，而是有很多的局部最优

3.用MSE作为二元分类的损失函数会有梯度消失的问题



**9.生成式模型和判别式模型的区别**

来源：映客直播 算法岗：[https://www.nowcoder.com/discuss/204240](https://www.nowcoder.com/discuss/204240" \t "_blank)

课程相关内容：监督学习入门-统计进阶+贝叶斯

回答：

假设有样本输入值（或者观察值）x，类别标签（或者输出值）y

判别式模型评估对象是最大化条件概率p(y|x)并直接对其建模，

生成式模型评估对象是最大化联合概率p(x,y)并对其建模。

**10.bagging vs boosting简述,Bagging能不能接其他的基模型,GBDT和RF的不同**

来源：字节跳动 算法岗：[https://www.nowcoder.com/discuss/205260](https://www.nowcoder.com/discuss/205260" \t "_blank)

来源：京东 算法岗：<https://www.nowcoder.com/discuss/199755>

来源：今日头条 算法岗<https://www.nowcoder.com/discuss/190742>

课程相关内容：更多监督学习算法-集成学习

回答：

Bagging:

主要是对数据进行有放回的采样,通过样本进行多个学习器的构建,最终结果由多个弱学习器以某种方式得到(比如取平均值或者进行投票)

Boosting:

对数据整体进行训练,对训练错误的样本提高权值,继续进行迭代训练.

就好像一个学生做错了题,然后抄在错题本上,以后经常注意学习一样(就相当于提高了错误的比重)

**11.梯度下降陷入局部最优有什么解决办法**

来源：TP-LINK 算法岗：[https://www.nowcoder.com/discuss/200260](https://www.nowcoder.com/discuss/200260" \t "_blank)

课程相关内容：监督学习入门-效果评估介绍（线性回归回顾）

回答：采用其他的优化算法，例如Momentum，Adagrad，RMSprop，Adam

**12.介绍一下PCA**

来源：摩根大通公司 算法岗：<https://www.nowcoder.com/discuss/198905>

课程相关内容：毕业实战设计-特征工程

回答：成分分析方法，是一种使用最广泛的数据降维算法。PCA的主要思想是将n维特征映射到k维上，这k维是全新的正交特征也被称为主成分，是在原有n维特征的基础上重新构造出来的k维特征。PCA的工作就是从原始的空间中顺序地找一组相互正交的坐标轴，新的坐标轴的选择与数据本身是密切相关的。其中，第一个新坐标轴选择是原始数据中方差最大的方向，第二个新坐标轴选取是与第一个坐标轴正交的平面中使得方差最大的，第三个轴是与第1,2个轴正交的平面中方差最大的。依次类推，可以得到n个这样的坐标轴。通过这种方式获得的新的坐标轴，我们发现，大部分方差都包含在前面k个坐标轴中，后面的坐标轴所含的方差几乎为0。于是，我们可以忽略余下的坐标轴，只保留前面k个含有绝大部分方差的坐标轴。事实上，这相当于只保留包含绝大部分方差的维度特征，而忽略包含方差几乎为0的特征维度，实现对数据特征的降维处理。

**13.softmax代码实现，梯度公式推导和代码实现，线性不可分的时候怎么办，反向传播的原理，RELU和Sigmoid相比，优点有哪些**

来源：TP-LINK 算法岗：<https://www.nowcoder.com/discuss/200260>

阿里妈妈-算法岗：<https://www.nowcoder.com/discuss/172988>

百度-算法岗：<https://www.nowcoder.com/discuss/198905>

课程相关内容：更多监督学习算法-神经网络算法

回答：...

1**4.FM(Factorization Machine,)，FFM（Field-aware Factorization Machines）**

[vivo-算法岗：https://www.nowcoder.com/discuss/197410](https://www.nowcoder.com/discuss/197410)

课程相关内容：P3

回答：因子分解机（Factorization Machine, FM）主要用于解决数据稀疏的业务场景（如推荐业务），特征怎样组合的问题

**15.偏差和方差区别**

百度-算法岗：<https://www.nowcoder.com/discuss/198393>

亚马逊Amazon-算法岗：<http://sofasofa.io/interviews/Default_open.php>

课程相关内容：更多监督学习算法-集成学习

回答：

**偏差**度量了学习算法的期望预测与真实结果的偏离程度，即刻画了算法本身的拟合能力。

**方差**度量了同样大小的训练集的变动所导致的学习性能变化，即刻画了数据扰动所造成的影响

1. Bias和模型复杂度的关系：当模型复杂度上升时，Bias减小。当模型复杂度降低时，Bias增加
2. Variance和模型复杂度的关系：当模型复杂度低时，Variance更低，当模型复杂度高时，Variance更高

**16.简述什么是凸优化**

字节跳动-算法岗：<http://sofasofa.io/interviews/Default_open.php>

课程相关内容：监督学习入门-效果评估介绍

回答：目标函数为凸函数、可行域为凸集的优化问题就是凸优化问题。

**17.二元分类问题经常会遇到非平衡数据的情况，也就是正样本和负样本的数量相差很大。在数据处理阶段，有哪些方法可以克服非平衡性？在评价模型时，可以使用哪些度量标准？**

谷歌 - 算法岗：<http://sofasofa.io/interviews/Default_open.php>

课程相关内容：监督学习入门-效果评估介绍

回答：在数据预处理阶段，为了解决非平衡性的问题，通常会通过采样的方法让两者数量尽量平衡。比如对数量较少的标签样本进行过采样(over-sampling)，或者对数量较多标签的样本进行欠采样(down-sampling)，也可以进行聚类抽样或者smote采样。

**18.决策树过拟合怎么办，决策树剪枝方法**

作业帮-算法岗 <https://www.nowcoder.com/discuss/182629>

课程相关内容: P1 机器学习初步 - 决策树

回答：

决策树的剪枝基本策略有 预剪枝 (Pre-Pruning) 和 后剪枝 (Post-Pruning)。

先对数据集划分成训练集和验证集，训练集用来决定树生成过程中每个结点划分所选择的属性；对于**预剪枝**，在每一次实际对结点进行进一步划分之前，先采用验证集的数据来验证如果划分是否能提高划分的准确性。如果不能，就把结点标记为叶结点并退出进一步划分；如果可以就继续递归生成节点。**后剪枝**则是先从训练集生成一颗完整的决策树，然后自底向上地对非叶结点进行考察，若将该结点对应的子树替换为叶结点能带来泛化性能提升，则将该子树替换为叶结点。

**19.假设我们有很多数据给用户看到，但是我们只知道一部分的数据用户是否点击了，其他的我们都不知道用户是否点击了（相当于没有标签），如果是你的话，你怎么解决这个问题？**

阿里巴巴 算法岗：<https://www.nowcoder.com/discuss/172988/>

课程相关内容：监督学习入门 - 逻辑回归

回答：...

1.什么是ROC曲线？ （监督学习入门II-模型效果评估指标）

回答：ROC曲线是Receiver Operating Characteristic Curve的简称，中文名为“受试者工作特征曲线”。

ROC曲线的横坐标为假阳性率（FPR）；纵坐标为真阳性率（TPR）。

2.ROC曲线相比P-R曲线有什么特点？（监督学习入门II-模型效果评估指标）

回答：相比P-R曲线，ROC曲线有一个特点，当正负样本的分布发生变化变化时，ROC曲线的形状能够基本保持不变，而P-R曲线的形状一半会发生剧烈的变化。

这个特点让ROC曲线能够尽量降低不同测试集带来的干扰，更加客观地衡量模型本身的性能。

3.在对模型进行过充分的离线评估之后，为什么还要进行在线A/B测试？（统计周-AB测试）

回答：

（1）离线评估无法完全消除模型过拟合的影响。

（2）离线评估无法完全欢迎线上的工程环境。离线评估往往不会考虑线上环境的延迟、数据丢失、标签数据缺失等情况。

（3）线上系统的某些商业指标在离线估计中无法计算。离线评估一般是针对模型本身进行评估，而与模型相关的其他指标，往往无法直接获得。

4.在模型评估过程中，过拟合和欠拟合具体是指什么现象？（监督学习入门II）

回答：过拟合是指模型对于训练数据拟合呈过当的情况，反映到评估指标上，就是模型在训练集上表现很好，但在测试集和新数据上表现较差。欠拟合指的是模型在训练和预测时表现都不好的情况。

5.降低过拟合和欠拟合的方法（监督学习入门II）

回答：

降低过拟合的方法

1. 获取更多的训练数据。
2. 降低模型复杂度。
3. 正则化方法。
4. 集成学习方法。

降低欠拟合的方法

1. 添加新特征。
2. 增加模型复杂度。
3. 减小正则化系数。