

蒋豆芽

机器学习——PCA与LDA 已购

来自【机器学习面试题汇总与解析（蒋豆芽面试题总结）】 | 74 浏览 | 0 回复 | 2021-06-25



蒋豆芽



+关注

机器学习面试题汇总与解析——PCA与LDA

1. PCA介绍一下☆☆☆☆☆
2. PCA算法步骤☆☆☆☆☆
3. PCA原理☆☆☆☆☆
4. PCA降维之后的维度怎么确定☆☆☆☆☆
5. 说说PCA的优缺点☆☆☆☆☆
6. 推导一下PCA☆☆☆☆☆
7. 降维方法有哪些？☆☆☆☆☆
8. LDA介绍一下☆☆☆☆☆
9. LDA的中心思想是什么☆☆☆☆☆
10. LDA的优缺点☆☆☆☆☆
11. 说说LDA的步骤☆☆☆☆☆
12. 推导一下LDA☆☆☆☆☆
13. PCA和LDA有什么区别☆☆☆☆☆
14. 偏差与方差☆☆☆☆☆
15. SVD懂么☆☆☆☆☆
16. 方差和协方差的理解☆☆☆☆☆
17. 伯努利分布和二项分布的区别☆☆☆☆☆

- =====
- 本专栏适合于Python已经入门的学生或人士，有一定的编程基础。
 - 本专栏适合于算法工程师、机器学习、图像处理求职的学生或人士。
 - 本专栏针对面试题答案进行了优化，尽量做到好记、言简意赅。这才是一份面试题总结的正确打开方式。这样才方便背诵
 - 如专栏内容有错漏，欢迎在评论区指出或私聊我更改，一起学习，共同进步。

蒋豆芽

关于**机器学习算法**书籍，我强烈推荐一本《**百面机器学习算法工程师带你面试**》，这个就很类似面经，还有讲解，写得比较好。私聊我进群。

关于**深度学习算法**书籍，我强烈推荐一本《**解析神经网络——深度学习实践手册**》，简称CNN book，通俗易懂。私聊我进群。

参考资料

B站机器学习视频: <https://space.bilibili.com/10781175/channel/detail?cid=133301>

PCA与LDA: 《百面机器学习算法工程师带你面试》第四章——降维

读者可以把参考文章看看

1. PCA介绍一下☆☆☆☆☆

参考回答

主成分分析 (Principal Component Analysis, PCA) 是一种多变量统计方法，它是最常用的降维方法之一，通过**正交变换**将一组可能存在**相关性**的变量数据转换为一组**线性不相关**的变量，转换后的变量被称为**主成分**。

可以使用两种方法进行 PCA，分别是特征分解或奇异值分解 (SVD)。PCA旨在找到数据中的**主成分**，并利用这些主成分表征原始数据，从而达到降维的目的。

算法步骤：

假设有m条n维数据。

1. 将原始数据按列组成n行m列矩阵X
2. 将X的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值
3. 求出协方差矩阵 $C = \frac{1}{m} X X^T$
4. 求出协方差矩阵的特征值以及对应的特征向量
5. 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k行组成矩阵P
6. $Y = PX$ 即为降维到k维后的数据

答案解析

PCA是比较常见的线性降维方法,通过**线性投影**将**高维数据映射到低维数据**中,所期望的是在投影的维度上,新特征自身的**方差**尽量大,方差越大特征越有效,尽量使产生的新特征间的**相关性**越小。

蒋豆芽

类似的问题还有：

2. 说说PCA的步骤☆☆☆☆☆

参考回答

算法步骤：

假设有m条n维数据。

1. 将原始数据按列组成n行m列矩阵 X
2. 将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值
3. 求出协方差矩阵 $C = \frac{1}{m} X X^T$
4. 求出协方差矩阵的特征值以及对应的特征向量
5. 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k行组成矩阵 P
6. $Y = PX$ 即为降维到k维后的数据

答案解析

无。

3. PCA原理☆☆☆☆☆

参考回答

PCA是比较常见的线性降维方法,通过**线性投影**将**高维数据映射到低维数据**中,所期望的是在投影的维度上,新特征自身的**方差**尽量大,方差越大特征越有效,尽量使产生的新特征间的**相关性**越小。

PCA算法的具体操作为对所有的样本进行**中心化操作**,计算样本的**协方差矩阵**,然后对协方差矩阵做**特征值分解**,取最大的n个特征值对应的特征向量构造**投影矩阵**。

答案解析

无。

4. PCA降维之后的维度怎么确定☆☆☆☆☆

参考回答

1. 可以利用交叉验证，再选择一个很简单的分类器，来选择比较好的 k 的值
2. 可以设置一个比重阈值 t，比如 95%，然后选择满足阈值的最小的 k：

蒋豆芽

$\sum_{i=1}^n \lambda_i$ 牛客@蒋豆芽

答案解析

无。

5. 说说PCA的优缺点☆☆☆☆☆

参考回答

优点

1. 仅仅需要以方差衡量信息量，不受数据集以外的因素影响
2. 各主成分之间正交，可消除原始数据成分间的相互影响的因素
3. 计算方法简单，主要运算是特征值分解，易于实现。

缺点

1. 主成分各个特征维度的含义具有一定的模糊性，不如原始样本特征的解释性强
2. 方差小的非主成分也可能含有对样本差异的重要信息，因此降维丢弃可能对后续数据处理有影响
3. PCA属于有损压缩

答案解析

无。

6. 推导一下PCA☆☆☆☆☆

参考回答

参考文章内容，自行推导

答案解析

无。

7. 降维方法有哪些？☆☆☆☆☆

参考回答

PCA：略

:三 蒋豆芽

答案解析

无。

8. 介绍一下LDA☆☆☆☆☆

参考回答

线性判别分析 (Linear Discriminant Analysis, LDA) 是一种基于**有监督学习**的**降维方式**,将数据集在低维度的空间进行投影,要使得投影后的**同类别**的数据点间的距离**尽可能的靠近**,而**不同类别**间的数据点的距离**尽可能的远**。

答案解析

无。

9. LDA的中心思想是什么☆☆☆☆☆

参考回答

最大化类间距离和最小化类内距离。

答案解析

无。

10. LDA的优缺点☆☆☆☆☆

参考回答

优点:

1. 在降维过程中可以使用类别的先验知识经验, 而像PCA这样的无监督学习则无法使用类别先验知识。
2. LDA在样本分类信息依赖均值而不是方差的时候, 比PCA之类的算法较优。

缺点

1. LDA不适合对非高斯分布样本进行降维, PCA也有这个问题。
2. LDA降维最多降到类别数 $k-1$ 的维数, 如果我们降维的维度大于 $k-1$, 则不能使用LDA。当然目前有一些LDA的进化版算法可以绕过这个问题。
3. LDA在样本分类信息依赖方差而不是均值的时候, 降维效果不好。
4. LDA可能过度拟合数据。

蒋豆芽

无。

11. 说说LDA的步骤☆☆☆☆☆

参考回答

算法步骤：

假设有m条n维数据。

1. 计算类内散度矩阵 S_w
2. 计算类间散度矩阵 S_b
3. 计算矩阵 $S_w^{-1}S_b$
4. 计算矩阵 $S_w^{-1}S_b$ 的最大的d个特征值和对应的d个特征向量 (w_1, w_2, \dots, w_d) ，得到投影矩阵 W
5. 对样本集中的每一个样本特征 x_i ，转化为新的样本 $z_i = W^T x_i$
6. 得到输出样本集 $D' = (z_1, y_1), (z_2, y_2), \dots, (z_m, y_m)$

答案解析

无。

12. 推导一下LDA☆☆☆☆☆

参考回答

参考文章内容，自行推导

答案解析

无。

13. PCA和LDA有什么区别☆☆☆☆☆

参考回答

LDA用于降维，和PCA有很多相同，也有很多不同的地方，因此值得好好的比较一下两者的降维异同点。

相同点：

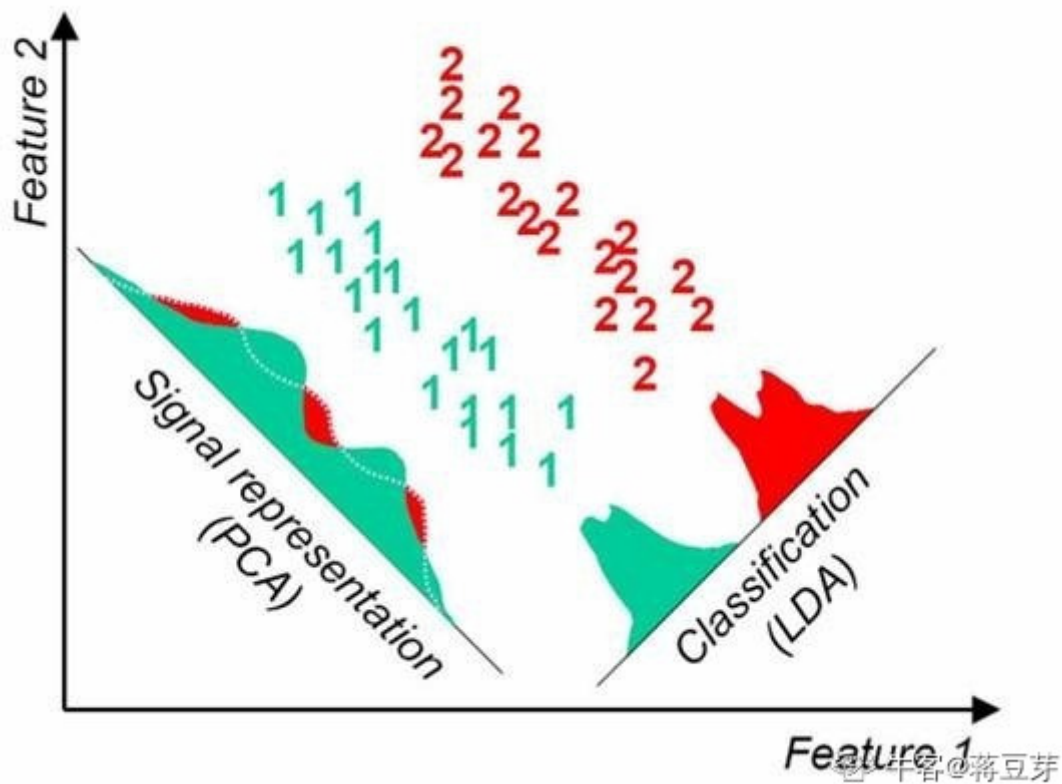
1. 两者均可以对数据进行降维。
2. 两者在降维时均使用了矩阵特征分解的思想。

蒋豆芽

1. LDA是**有监督**的降维方法，而PCA是**无监督**的降维方法
2. LDA降维最多降到类别数 $k-1$ 的维数，而PCA没有这个限制。
3. LDA除了可以用于**降维**，还可以用于**分类**。
4. L***择分类性能最好的投影方向**，而PCA**选择样本点投影具有最大方差的方向****。

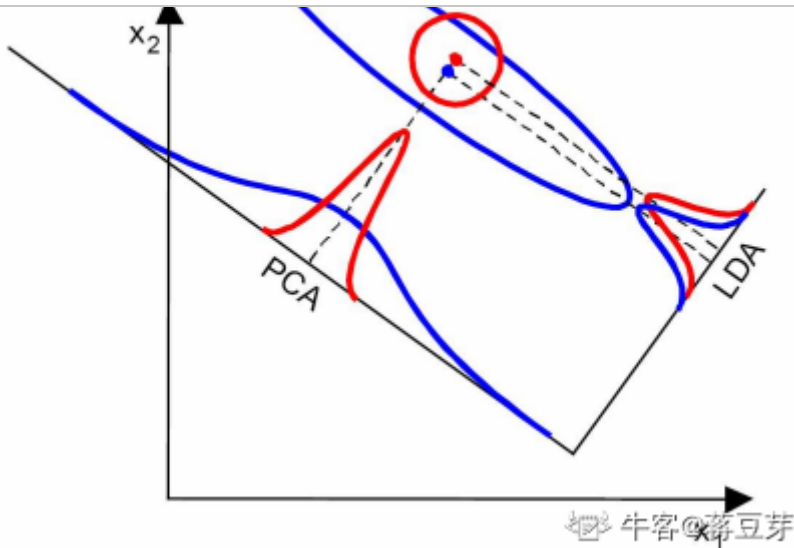
答案解析

这点可以从下图形象的看出，在某些数据分布下LDA比PCA降维较优。



当然，某些某些数据分布下PCA比LDA降维较优，如下图所示：

蒋豆芽



14. 偏差与方差☆☆☆☆

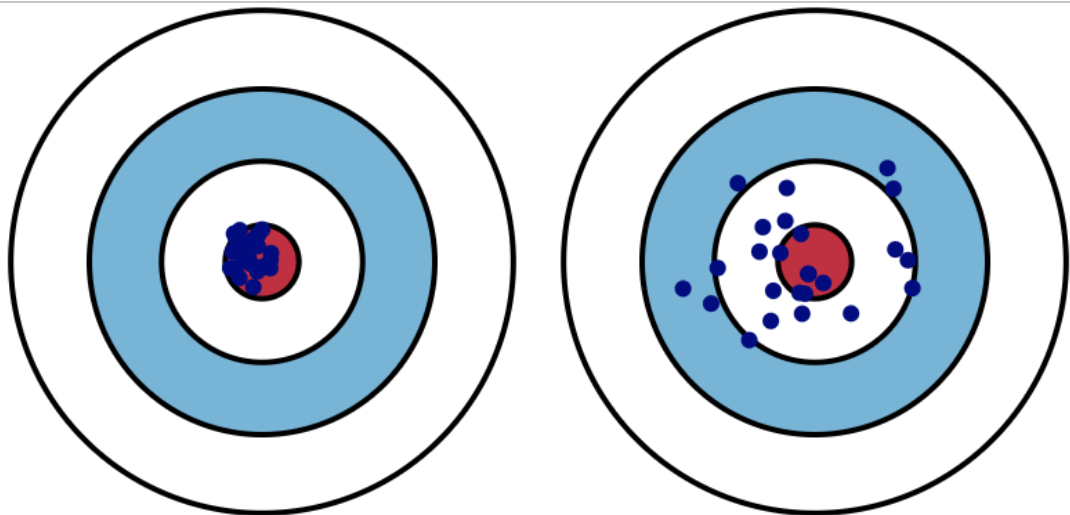
参考回答

偏差 (bias)： 偏差衡量了模型的**预测值与实际值之间的偏离关系**。通常在深度学习中，我们每一次训练迭代出来的新模型，都会拿训练数据进行预测，偏差就反应在预测值与实际值匹配度上，比如通常在keras运行中看到的准确度为96%，则说明是低偏差；反之，如果准确度只有70%，则说明是高偏差。

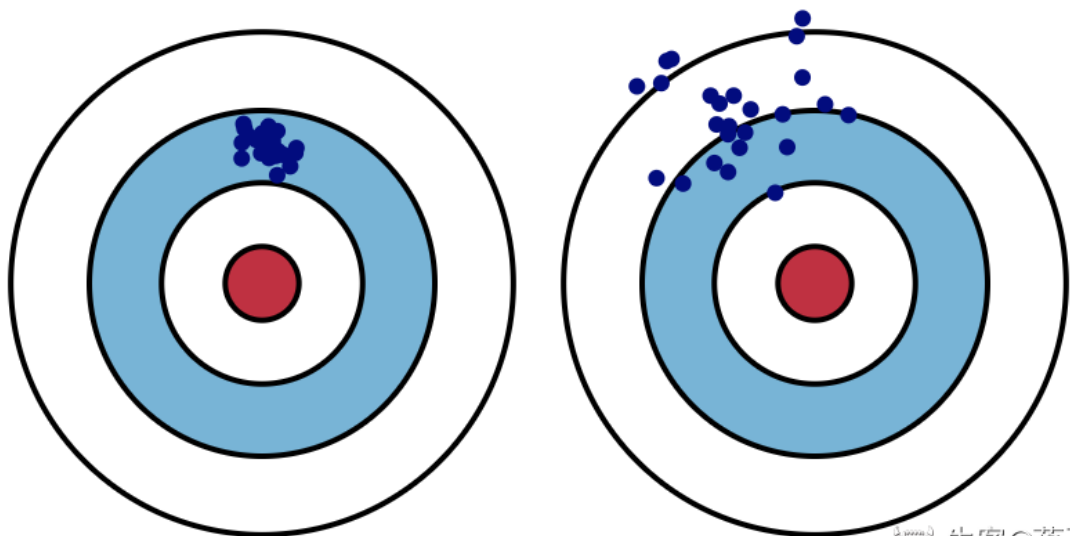
方差 (variance)： 方差描述的是训练数据在不同迭代阶段的训练模型中，预测值的变化波动情况（或称之为离散情况）。从数学角度看，可以理解为每个**预测值与预测均值差的平方和的再求平均数**。通常在深度学习训练中，初始阶段模型复杂度不高，为低方差；随着训练量加大，模型逐步拟合训练数据，复杂度开始变高，此时方差会逐渐变高。

答案解析

Low Bias



High Bias



牛客@蒋豆芽

这是一张常见的靶心图。可以想象红色靶心表示为实际值，蓝色点集为预测值。在模型不断地训练迭代过程中，我们能碰到四种情况：

1. **低偏差，低方差**：这是训练的理想模型，此时蓝色点集基本落在靶心范围内，且数据离散程度小，基本在靶心范围内；
2. **低偏差，高方差**：这是深度学习面临的最大问题，过拟合了。也就是模型太贴合训练数据了，导致其泛化（或通用）能力差，若遇到测试集，则准确度下降的厉害；
3. **高偏差，低方差**：这往往是训练的初始阶段；
4. **高偏差，高方差**：这是训练最糟糕的情况，准确度差，数据的离散程度也差。

15. SVD懂么☆☆☆☆☆

参考回答

奇异值分解(Singular Value Decomposition, 以下简称SVD)是在机器学习领域广泛应用的算法，它不光可以用于降维算法中的特征分解，还可以用于推荐系统，以及自然语言处理等领域。

蒋豆芽

$m \times m$ 的矩阵, Σ 是一个 $m \times n$ 的矩阵, V 是一个 $n \times n$ 的矩阵, $U^T U = I, V^T V = I$ 。那么 AA^T 的特征向量组成的就是我们SVD中的 U 矩阵。

答案解析

无。

16. 方差和协方差的理解 ☆ ☆ ☆ ☆

参考回答

方差：度量单个随机变量的离散程度，公式如下：

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

牛客@蒋豆芽

协方差：度量两个随机变量（变化趋势）的相似程度，定义如下：

$$\sigma(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

牛客@蒋豆芽

答案解析

无。

17. 伯努利分布和二项分布的区别 ☆ ☆ ☆ ☆


参考回答

伯努利分布：是假设一个事件只有发生或者不发生两种可能，并且这两种可能是固定不变的。那么，如果假设它发生的概率是 p ，那么它不发生的概率就是 $1-p$ 。这就是伯努利分布。

二项分布是多次伯努利分布实验的**概率分布**。

答案解析

无。

 蒋豆芽[资源分享](#)[python](#)[机器学习](#)[算法工程师](#)[春秋招](#)[面试题](#)[软件开发](#)[国经](#)[牛客](#)

收藏



赞

相关专栏



机器学习面试题汇总与解析（蒋豆芽面试题总结）

27篇文章 | 90订阅

[已订阅](#)

0条评论

[默认排序](#) 

没有回复

请留下你的观点吧~

[发布](#) **牛客博客，记录你的成长**[关于博客](#) | [意见反馈](#) | [免责声明](#) | [牛客网首页](#)