

 蒋豆芽

机器学习——SVM 已购

来自 [【机器学习面试题汇总与解析（蒋豆芽面试题总结）】](#) | 78 浏览 | 0 回复 | 2021-05-22



蒋豆芽

[+关注](#)

机器学习面试题汇总与解析——SVM

1. 推导SVM☆☆☆☆☆
2. LR 和 SVM 联系与区别☆☆☆☆☆
3. svm介绍一下☆☆☆☆☆
4. 讲一下SVM的原理☆☆☆☆☆
5. 如果特征比较多，用LR还是SVM?☆☆☆☆☆
6. 介绍SVM☆☆☆☆☆
7. SVM是否可以用随机梯度下降☆☆☆☆☆
8. SVM优缺点☆☆☆☆☆
9. 为什么要将求解 SVM 的原始问题转换为其对偶问题☆☆☆☆☆
10. 为什么SVM对缺失数据敏感☆☆☆☆☆
11. SVM怎么防止过拟合?☆☆☆☆☆

- =====
- 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。
 - 相信大家都有着高尚的灵魂，请尊重我的知识产权，未经允许严禁各类机构和个人转载、传阅本专栏的内容。
- =====

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。

蒋豆芽

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

SVM: <https://www.cnblogs.com/vpegasus/p/svm.html> (这篇基本讲清楚了SVM, 一定要仔细看)

SVM实战: <https://www.cnblogs.com/zy230530/p/6901277.html>

读者可以把参考文章看看

个人理解

自己动手推推公式就明白了。

SVM的优缺点

优点:

1. 理论完善, 逻辑优美。
2. SVM 的最终决策函数**只由少数的支持向量所确定**, 计算的复杂性取决于**支持向量的数目**, 而不是样本空间的**维数**, 这在某种意义上避免了“维数灾难”。
3. SVM 对异常值不敏感
4. 少数支持向量决定了最终结果, 这不但可以帮助我们抓住关键样本、“剔除”大量冗余样本
5. SVM 目标是求解最大间隔超平面, 算法分类具有较好的“鲁棒”性。
6. SVM可以利用核函数解决逻辑回归解决不了的高维分类问题。

缺点:

1. **SVM**算法对大规模训练样本难以实施

由于**SVM**是借助二次规划来求解支持向量, 而求解二次规划将涉及 m 阶矩阵的计算 (m 为样本的个数), 当 m 数目很大时该矩阵的存储和计算将耗费大量的机器内存和运算时间。

2. 用**SVM**解决多分类问题存在困难

经典的支持向量机算法只给出了二类分类的算法, 而在数据挖掘的实际应用中, 一般要解决多类的分类问题。可以通过多个二类支持向量机的组合来解决。主要有一对多组合模式、一对一组合模式和SVM决策树; 再就是通过构造多个分类器的组合来解决。主要原理是克服SVM固有的缺点, 结合其他算法的优势, 解决多类问题的分类精度。如: 与粗集理论结合, 形成一种优势互补的多类问题的组合分类器。

1. 推导SVM☆☆☆☆☆

参考回答

蒋豆芽

答案解析

无。

类似的问题还有：

2. LR 和 SVM 联系与区别 ☆ ☆ ☆ ☆ ☆

参考回答

LR与SVM的相同点：

1. 都是有监督的分类算法；
2. 如果不考虑核函数，LR和SVM都是线性分类算法。

它们的分类决策面都是线性的。

3. LR和SVM都是判别式模型。

LR与SVM的不同点：

1. 本质上是loss函数不同，或者说分类的原理不同。
2. SVM是结构风险最小化，LR则是经验风险最小化。
3. SVM只考虑分界面附近的少数点，而LR则考虑所有点。
4. 在解决非线性问题时，SVM可采用核函数的机制，而LR通常不采用核函数的方法。
5. SVM计算复杂，但效果比LR好，适合小数据集；LR计算简单，适合大数据集，可以在线训练。

答案解析

判别式模型 (Discriminative Model)：直接对条件概率 $p(y|x)$ 进行建模，常见判别模型有：线性回归、决策树、支持向量机SVM、k近邻、神经网络等；

生成式模型 (Generative Model)：对联合分布概率 $p(x,y)$ 进行建模，常见生成式模型有：隐马尔可夫模型HMM、朴素贝叶斯模型、高斯混合模型GMM、LDA等；

1. 生成式模型更普适；判别式模型更直接，目标性更强
2. 生成式模型关注数据是如何产生的，寻找的是数据分布模型；判别式模型关注的数据的差异性，寻找的是分类面
3. 由生成式模型可以产生判别式模型，但是由判别式模型没法形成生成式模型

3. svm介绍一下 ☆ ☆ ☆ ☆ ☆

参考回答

:三 蒋豆芽

型使用**拉格朗日方法**找到其对偶问题，从而找到解。**SVM**只和**支持向量**有关，所以**SVM**对**异常值**不敏感，适合**小数据集**。

对于不易分类的情况，可以采用**软间隔**。高维问题甚至可以采用**核函数**方法来更好的分类。

答案解析

无。

4. 讲一下SVM的原理☆☆☆☆☆

参考回答

SVM (Support Vector Machine, 支持向量机) 是二十世纪初使用广泛的分类算法，可以处理非线性和高维的机器学习问题。**SVM**目标是找到分类的**最大间隔**，为了更好的求解，针对**SVM**基本型使用**拉格朗日方法**找到其对偶问题，从而找到解。**SVM**只和**支持向量**有关，所以**SVM**对**异常值**不敏感，适合**小数据集**。

对于不易分类的情况，可以采用**软间隔**。高维问题甚至可以采用**核函数**方法来更好的分类。

答案解析

无。

5. 如果特征比较多，用LR还是SVM?☆☆☆☆☆

参考回答

如果特征比较多，最好采用**SVM**。因为从**SVM**基本型可以看出来，只和**样本数量**有关，和**特征**无关，不易过拟合；而特征过多时，**LR**容易过拟合。

答案解析

无。

6. 介绍SVM☆☆☆☆☆

参考回答

答案参考上面。

答案解析

无。

蒋豆芽

SVM本质上是一个带约束的二次规划问题，但是通过拉格朗日法或Hinge Loss的方式可以转换成一个**无约束**的优化问题。而不论是**梯度下降**还是**二次规划**，都是能够用来优化这个问题的。甚至Hinge Loss定义下的损失函数是个**凸函数**，是可以优化得到最优解的。

答案解析

无

8. SVM优缺点☆☆☆☆☆

参考回答

SVM的优缺点

优点：

1. 理论完善，逻辑优美。
2. SVM 的最终决策函数**只由少数的支持向量所确定**，计算的复杂性取决于**支持向量的数目**，而不是样本空间的**维数**，这在某种意义上避免了“维数灾难”。
3. SVM 对异常值不敏感
4. 少数支持向量决定了最终结果，这不但可以帮助我们抓住关键样本、“剔除”大量冗余样本
5. SVM 目标是求解最大间隔超平面，算法分类具有较好的“鲁棒”性。
6. SVM可以利用核函数解决逻辑回归解决不了的高维分类问题。

缺点：

1. **SVM**算法对大规模训练样本难以实施

由于**SVM**是借助二次规划来求解支持向量，而求解二次规划将涉及m阶矩阵的计算（m为样本的个数），当m数目很大时该矩阵的存储和计算将耗费大量的机器内存和运算时间。

2. 用**SVM**解决多分类问题存在困难

经典的支持向量机算法只给出了二类分类的算法，而在数据挖掘的实际应用中，一般要解决多类的分类问题。可以通过多个二类支持向量机的组合来解决。主要有一对多组合模式、一对一组合模式和SVM决策树；再就是通过构造多个分类器的组合来解决。主要原理是克服SVM固有的缺点，结合其他算法的优势，解决多类问题的分类精度。如：与粗集理论结合，形成一种优势互补的多类问题的组合分类器。

答案解析

无

9. 为什么要将求解 SVM 的原始问题转换为其对偶问题☆☆☆☆☆

蒋豆芽

搜寻的范围，但是却使问题变得更加复杂。为了使问题变得易于处理，我们的方法是把**目标函数**和**约束**全部融入一个新的函数，即**拉格朗日函数**，再通过这个函数来寻找最优点。二是可以自然引入**核函数**，进而推广到非线性分类问题。

答案解析

无。

10. 为什么SVM对缺失数据敏感☆☆☆☆☆

参考回答

这里说的**缺失数据**是指缺失某些**特征数据**，向量数据**不完整**。**SVM** 没有处理缺失值的策略。而 **SVM** 希望样本在特征空间中线性可分，所以特征空间的好坏对SVM的性能很重要。**缺失特征数据** 将影响训练结果的好坏。

答案解析

无。

11. SVM怎么防止过拟合？☆☆☆☆☆

参考回答

引入松弛变量

答案解析

无。

[资源分享](#)[python](#)[机器学习](#)[算法工程师](#)[春秋招](#)[面试题](#)[软件开发](#)[面经](#)[举报](#)

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

 蒋豆芽



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](#)