**SVM常见面试题**

1. SVM 原理

SVM 是一种二类分类模型。它的基本模型是在特征空间中寻找间隔最大化的分离超平面的线性分类器。

* 当训练样本线性可分时，通过硬间隔最大化，学习一个线性分类器，即线性可分支持向量机；
* 当训练数据近似线性可分时，引入松弛变量，通过软间隔最大化，学习一个线性分类器，即线性支持向量机；
* 当训练数据线性不可分时，通过使用核技巧及软间隔最大化，学习非线性支持向量机。

以上各种情况下的数学推到应当掌握，硬间隔最大化（几何间隔）、学习的对偶问题、软间隔最大化（引入松弛变量）、非线性支持向量机（核技巧）。

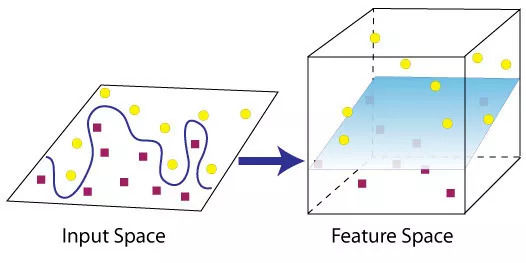
2. SVM 为什么采用间隔最大化

当训练数据线性可分时，存在无穷个分离超平面可以将两类数据正确分开。感知机利用误分类最小策略，求得分离超平面，不过此时的解有无穷多个。线性可分支持向量机利用间隔最大化求得最优分离超平面，这时，解是唯一的。另一方面，此时的分隔超平面所产生的分类结果是最鲁棒的，对未知实例的泛化能力最强。可以借此机会阐述一下几何间隔以及函数间隔的关系。

3. 为什么要将求解 SVM 的原始问题转换为其对偶问题

一是对偶问题往往更易求解，当我们寻找约束存在时的最优点的时候，约束的存在虽然减小了需要搜寻的范围，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。

4. 为什么 SVM 要引入核函数



当样本在原始空间线性不可分时，可将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分。而引入这样的映射后，所要求解的对偶问题的求解中，无需求解真正的映射函数，而只需要知道其核函数。核函数的定义：K(x,y)=<ϕ(x),ϕ(y)>，即在特征空间的内积等于它们在原始样本空间中通过核函数 K 计算的结果。一方面数据变成了高维空间中线性可分的数据，另一方面不需要求解具体的映射函数，只需要给定具体的核函数即可，这样使得求解的难度大大降低。

5. 为什么SVM对缺失数据敏感

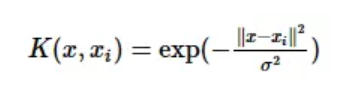
这里说的缺失数据是指缺失某些特征数据，向量数据不完整。SVM 没有处理缺失值的策略。而 SVM 希望样本在特征空间中线性可分，所以特征空间的好坏对SVM的性能很重要。缺失特征数据将影响训练结果的好坏。

6. SVM 核函数之间的区别

一般选择线性核和高斯核，也就是线性核与 RBF 核。 线性核：主要用于线性可分的情形，参数少，速度快，对于一般数据，分类效果已经很理想了。 RBF 核：主要用于线性不可分的情形，参数多，分类结果非常依赖于参数。有很多人是通过训练数据的交叉验证来寻找合适的参数，不过这个过程比较耗时。 如果 Feature 的数量很大，跟样本数量差不多，这时候选用线性核的 SVM。 如果 Feature 的数量比较小，样本数量一般，不算大也不算小，选用高斯核的 SVM。

以上是几个问题在面试中遇到 SVM 算法时，几乎是必问的问题，另外，大家一定要做到自己可以推导集合间隔、函数间隔以及对偶函数，并且理解对偶函数的引入对计算带来的优势。

7. svm RBF核函数的具体公式？



Gauss径向基函数则是局部性强的核函数，其外推能力随着参数σ的增大而减弱。

这个核会将原始空间映射为无穷维空间。不过，如果σ选得很大的话，高次特征上的权重实际上衰减得非常快，所以实际上（数值上近似一下）相当于一个低维的子空间；反过来，如果σ选得很小，则可以将任意的数据映射为线性可分——当然，这并不一定是好事，因为随之而来的可能是非常严重的过拟合问题。不过，总的来说，通过调控参数σ，高斯核实际上具有相当高的灵活性，也是使用最广泛的核函数之一。

8. SVM如何处理多分类问题？

一般有两种做法：一种是直接法，直接在目标函数上修改，将多个分类面的参数求解合并到一个最优化问题里面。看似简单但是计算量却非常的大。

另外一种做法是间接法：对训练器进行组合。其中比较典型的有一对一，和一对多。

一对多，就是对每个类都训练出一个分类器，由svm是二分类，所以将此而分类器的两类设定为目标类为一类，其余类为另外一类。这样针对k个类可以训练出k个分类器，当有一个新的样本来的时候，用这k个分类器来测试，那个分类器的概率高，那么这个样本就属于哪一类。这种方法效果不太好，bias比较高。

svm一对一法（one-vs-one），针对任意两个类训练出一个分类器，如果有k类，一共训练出C(2,k) 个分类器，这样当有一个新的样本要来的时候，用这C(2,k) 个分类器来测试，每当被判定属于某一类的时候，该类就加一，最后票数最多的类别被认定为该样本的类。

9.SVM是用的是哪个库？Sklearn/libsvm中的SVM都有什么参数可以调节？

用的是sklearn实现的。采用sklearn.svm.SVC设置的参数。本身这个函数也是基于libsvm实现的（PS: libsvm中的二次规划问题的解决算法是SMO）。

SVC函数的训练时间是随训练样本平方级增长，所以不适合超过10000的样本。

对于多分类问题，SVC采用的是one-vs-one投票机制，需要两两类别建立分类器，训练时间可能比较长。

sklearn.svm.SVC(C=1.0, kernel='rbf', degree=3, gamma='auto', coef0=0.0, shrinking=True, probability=False,tol=0.001, cache\_size=200, class\_weight=None, verbose=False, max\_iter=-1, decision\_function\_shape=None,random\_state=None)

参数：

l  C：C-SVC的惩罚参数C?默认值是1.0

C越大，相当于惩罚松弛变量，希望松弛变量接近0，即对误分类的惩罚增大，趋向于对训练集全分对的情况，这样对训练集测试时准确率很高，但泛化能力弱。C值小，对误分类的惩罚减小，允许容错，将他们当成噪声点，泛化能力较强。

l  kernel ：核函数，默认是rbf，可以是‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’

  　　0 – 线性：u'v

 　　 1 – 多项式：(gamma\*u'\*v + coef0)^degree

  　　2 – RBF函数：exp(-gamma|u-v|^2)

  　　3 –sigmoid：tanh(gamma\*u'\*v + coef0)

l  degree ：多项式poly函数的维度，默认是3，选择其他核函数时会被忽略。

l  gamma ： ‘rbf’,‘poly’ 和‘sigmoid’的核函数参数。默认是’auto’，则会选择1/n\_features

l  coef0 ：核函数的常数项。对于‘poly’和 ‘sigmoid’有用。

l  probability ：是否采用概率估计？.默认为False

l  shrinking ：是否采用shrinking heuristic方法，默认为true

l  tol ：停止训练的误差值大小，默认为1e-3

l  cache\_size ：核函数cache缓存大小，默认为200

l  class\_weight ：类别的权重，字典形式传递。设置第几类的参数C为weight\*C(C-SVC中的C)

l  verbose ：允许冗余输出？

l  max\_iter ：最大迭代次数。-1为无限制。

l  decision\_function\_shape ：‘ovo’, ‘ovr’ or None, default=None3

l  random\_state ：数据洗牌时的种子值，int值

10.SVM的有点

（1）非线性映射-理论基础 （2）最大化分类边界-方法核心 （3）支持向量-计算结果 （4）小样本学习方法 （5）最终的决策函数只有少量支持向量决定，避免了“维数灾难” （6）少数支持向量决定最终结果—->可“剔除”大量冗余样本+算法简单+具有鲁棒性（体现在3个方面） （7）学习问题可表示为凸优化问题—->全局最小值 （8）可自动通过最大化边界控制模型，但需要用户指定核函数类型和引入松弛变量 （9）适合于小样本，优秀泛化能力（因为结构风险最小） （10）泛化错误率低，分类速度快，结果易解释

11.SVM的缺点

优点：见主要特点

缺点：（1）大规模训练样本（m阶矩阵计算） （2）传统的不适合多分类 （3）对缺失数据、参数、核函数敏感

12.如何选择核函数

（1）线性可分 （2）线性不可分（多项式~，高斯~，拉普拉斯~，sigmoid~）

13.RBF核的优点

大小高低都适用。具体来说（1）无穷维，线性核是其特例 （2）与多项式~比，RBF需确定的参数少 （3）某些参数下，与sigmoid~有相似的功能

14.核函数选取与feature、样本之间的关系

（1）fea大≈样本数量：LR or 线性核 （2）fea小，样本数量不大也不小：高斯核 （3）fea大，样本数量多：手工添加特征后转（1）

15.介绍你所知道的几种核函数

多项式~，高斯~，拉普拉斯~，sigmoid~；线性~

16.如何调节惩罚因子

惩罚因子C表示有多重视离群点带来的损失，当所有离群点的松弛变量和一定是，C越大，对目标函数的损失也就越大。

C不是一个变量，是需要参数寻优的常量。

17.如何防止SVM过拟合（提高泛化能力）

松弛变量的平方和？

18.SVM与LR的区别与联系

联系：（1）分类（二分类） （2）可加入正则化项

区别：（1）LR–参数模型；SVM–非参数模型？（2）目标函数：LR—logistical loss；SVM–hinge loss （3）SVM–support vectors；LR–减少较远点的权重 （4）LR–模型简单，好理解，精度低，可能局部最优；SVM–理解、优化复杂，精度高，全局最优，转化为对偶问题—>简化模型和计算 （5）LR可以做的SVM可以做（线性可分），SVM能做的LR不一定能做（线性不可分）

区别整理：

1、样本点对模型的作用不同。SVM中，只有关键的样本点（支持向量）对模型结果有影响，而LR中，每一个样本点都对模型有影响。

2、损失函数不同。SVM是hinge损失函数，LR是log损失函数

3、理论基础不同。SVM基于严格的数学推导，LR基于统计。

4、输出不同。LR可以对每个样本点给出类别判断的概率值，SVM无法做到。

5、可处理的特征空间维度不同。LR在特征空间维度很高时，表现较差。SVM则可以通过对偶求解高效应对这一挑战。

6、防过拟合能力不同。SVM模型中内含了L2正则，可有效防止过拟合。LR要自己添加正则项。

7、处理非线性分类问题能力不同。SVM可通过核函数灵活地将非线性问题转化为线性分类问题。LR如果要做到这一点，需要自己手动地进行特征转换。

8、处理分类问题能力不同。SVM只能处理二类分类问题，如果要处理多类别分类，需要进行 one VS one 或one VS all建模。LR可以直接进行多类别分类。

9、计算复杂度不同。对于海量数据，SVM的效率较低，LR效率比较高。

10、对数据要求不同。SVM依赖于数据表达出的距离测度，所以需要对数据进行标准化处理，而LR不需要。

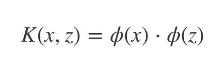
11、能力范围不同。 SVM拓展后，可解决回归问题，LR不能。

12、可解释性不同。LR基于统计，可解释性比SVM好。

13、抗噪声数据能力不同。SVM的损失函数基于距离测度，抗噪声能力要强于LR。

# 19.核技巧：

针对非线性的问题，如果能将样本从原始空间映射到更高维度的特征空间之上，在新的特征空间上样本就可能是线性可分的。如果样本的属性数有限，那么一定存在一个高维特征空间使样本可分。将原始低维空间上的非线性问题转化为新的高维空间上的线性问题，这就是核技巧的基本思想。

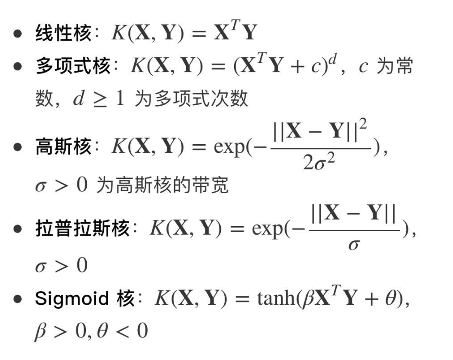
核技巧应用到支持向量机中时，原始空间与新空间之间的转化是通过非线性变换实现的。假设原始空间是低维欧几里得空间X，新空间为高维希尔伯特空间H，则从 X到 H的映射可以用函数   
https://diycode.b0.upaiyun.com/photo/2018/a053e6d43189ae92d13fa566529edaa4.png  
核函数可以表示成映射函数内积的形式：  


# 20.核函数有两个特点：

1. 其计算过程是在低维空间上完成的，因而避免了高维空间中复杂的计算；
2. 对于给定的核函数，高维空间 H和映射函数 ϕ的取法并不唯一。而且一方面，高维空间的取法可以不同；另一方面，即使在同一个空间上，映射函数也可以有所区别。

一般的核函数都是正定核函数。正定核函数的充要条件是由函数中任意数据的集合形成的核矩阵都是半正定的，这意味着任何一个核函数都隐式定义了一个成为“再生核希尔伯特空间”的特征空间。不好的核函数会将样本映射到不合适的特征空间，导致分类效果不好。

讲到这里大家应该猜出来了，核函数的选择就是svm中的难点，也是核心问题。

常用的核函数有：  


为什么叫“支持向量机”，是因为当训练完成后，最终模型只与支持向量有关，而它的关键是如何根据支持向量构建出解，算法的复杂度也主要取决于支持向量的数目。