

1. 关于支持向量机 SVM, 下列说法错误的是 ()

- A. L2 正则项, 作用是最大化分类间隔, 使得分类器拥有更强的泛化能力
- B. Hinge 损失函数, 作用是最小化经验分类错误
- C. 分类间隔为, $\frac{1}{\|w\|}$ 代表向量的模
- D. 当参数 C 越小时, 分类间隔越大, 分类错误越多, 趋于欠学习

正确答案: C

解析:

A 正确。考虑加入正则化项的原因: 想象一个完美的数据集, $y > 1$ 是正类, $y < -1$ 是负类, 决策面 $y = 0$, 加入一个 $y = -30$ 的正类噪声样本, 那么决策面将会变“歪”很多, 分类间隔变小, 泛化能力减小。加入正则项之后, 对噪声样本的容错能力增强, 前面提到的例子里面, 决策面就会没那么“歪”了, 使得分类间隔变大, 提高了泛化能力。

B 正确。

C 错误。间隔应该是 $\frac{1}{\|w\|}$ 才对, 后半句应该没错, 向量的模通常指的就是其二范数。

D 正确。考虑软间隔的时候, C 对优化问题的影响就在于把 a 的范围从 $[0, +\infty]$ 限制到了 $[0, C]$ 。C 越小, 那么 a 就会越小, 目标函数拉格朗日函数导数为 0 可以求出, a 变小使得 w 变小, 因此间隔变大

2. 关于 Logit 回归和 SVM 不正确的是 ()

- A. Logit 回归本质上是一种根据样本对权值进行极大似然估计的方法, 而后验概率正比于先验概率和似然函数的乘积。logit 仅仅是最大化似然函数, 并没有最大化后验概率, 更谈不上最小化后验概率。A 错误
- B. Logit 回归的输出就是样本属于正类别的几率, 可以计算出概率, 正确
- C. SVM 的目标是找到使得训练数据尽可能分开且分类间隔最大的超平面, 应该属于结构风险最小化。
- D. SVM 可以通过正则化系数控制模型的复杂度, 避免过拟合。

正确答案: A

解析:

Logit 回归目标函数是最小化后验概率, Logit 回归可以用于预测事件发生概率的大小, SVM 目标是结构风险最小化, SVM 可以有效避免模型过拟合。

3. 在其他条件不变的前提下, 以下哪种做法容易引起机器学习中的过拟合问题 ()

- A. 增加训练集量
- B. 减少神经网络隐藏层节点数
- C. 删除稀疏的特征
- D. SVM 算法中使用高斯核/RBF 核代替线性核

正确答案: (D)

解析:

一般情况下, 越复杂的系统, 过拟合的可能性就越高, 一般模型相对简单的话泛化能力会更好一点。

B.一般认为，增加隐层数可以降低网络误差（也有文献认为不一定能有效降低），提高精度，但也使网络复杂化，从而增加了网络的训练时间和出现“过拟合”的倾向，svm 高斯核函数比线性核函数模型更复杂，容易过拟合

D.径向基(RBF)核函数/高斯核函数的说明.这个核函数可以将原始空间映射到无穷维空间。对于参数，如果选的很大，高次特征上的权重实际上衰减得非常快，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，通过调整参数，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。

4. 关于 Logit 回归和 SVM 不正确的是（ ）。

- A. Logit 回归目标函数是最小化后验概率
- B. Logit 回归可以用于预测事件发生概率的大小
- C. SVM 目标是结构风险最小化
- D. SVM 可以有效避免模型过拟合

正确答案：（A）

解析：

A. Logit 回归本质上是一种根据样本对权值进行极大似然估计的方法，而后验概率正比于先验概率和似然函数的乘积。logit 仅仅是最大化似然函数，并没有最大化后验概率，更谈不上最小化后验概率。而最小化后验概率是朴素贝叶斯算法要做的。A 错误

B. Logit 回归的输出就是样本属于正类别的几率，可以计算出概率，正确

C. SVM 的目标是找到使得训练数据尽可能分开且分类间隔最大的超平面，应该属于结构风险最小化。

D. SVM 可以通过正则化系数控制模型的复杂度，避免过拟合。

5. 有两个样本点，第一个点为正样本,它的特征向量是(0,-1);第二个点为负样本,它的特征向量是(2,3),从这两个样本点组成的训练集构建一个线性 SVM 分类器的分类面方程是（ ）

- A. $2x+y=4$
- B. $x+2y=5$
- C. $x+2y=3$
- D. $2x-y=0$

正确答案：（C）

解析：

这道题简化了，对于两个点来说，最大间隔就是垂直平分线，因此求出垂直平分线即可。

6. 关于支持向量机 SVM,下列说法错误的是（ ）

- A. L2 正则项，作用是最大化分类间隔，使得分类器拥有更强的泛化能力
- B. Hinge 损失函数，作用是最小化经验分类错误
- C. 分类间隔为 $1/\|w\|$, $\|w\|$ 代表向量的模
- D. 当参数 C 越小时，分类间隔越大，分类错误越多，趋于欠学习

正确答案：（C）

解析：

A 正确。考虑加入正则化项的原因：想象一个完美的数据集， $y>1$ 是正类， $y<-1$ 是负类，决策面 $y=0$ ，加入一个 $y=-30$ 的正类噪声样本，那么决策面将会变“歪”很多，分类间隔变小，泛化能力减小。加入正则项之后，对噪声样本的容错能力增强，前面提到的例子里面，决策面就会没那么“歪”了，使得分类间隔变大，提高了泛化能力。

B 正确。

C 错误。间隔应该是 C 才对，后半句应该没错，向量的模通常指的就是其二范数。

D 正确。考虑软间隔的时候， C 对优化问题的影响就在于把 a 的范围从 $[0, +\infty]$ 限制到了 $[0, C]$ 。 C 越小，那么 a 就会越小，目标函数拉格朗日函数导数为 0 可以求出， a 变小使得 w 变小，因此间隔变大。

7. 下列不是 SVM 核函数的是：（ ）

- A. 多项式核函数
- B. Logistic 核函数
- C. 径向基核函数
- D. Sigmoid 核函数

正确答案：（B）

解析：

SVM 核函数包括线性核函数、多项式核函数、径向基核函数、高斯核函数、幂指数核函数、拉普拉斯核函数、ANOVA 核函数、二次有理核函数、多元二次核函数、逆多元二次核函数以及 Sigmoid 核函数。核函数的定义并不困难，根据泛函的有关理论，只要一种函数 $K(x_i, x_j)$ 满足 Mercer 条件，它就对应某一变换空间的内积。对于判断哪些函数是核函数到目前为止也取得了重要的突破，得到 Mercer 定理和以下常用的核函数类型：

(1) 线性核函数： $K(x, x_i) = x \cdot x_i$

(2) 多项式核： $K(x, x_i) = ((x \cdot x_i) + 1)^d$

(3) 径向基核 (RBF)： $K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$

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。多项式形式的核函数具有良好的全局性质。局部性较差。

(4) 傅里叶核： $K(x, x_i) = 1 - q^2 (1 - 2q \cos(x - x_i) + q^2)$

(5) 样条核： $K(x, x_i) = B_{2n+1}(x - x_i)$

(6) Sigmoid 核函数： $K(x, x_i) = \tanh(\kappa(x, x_i) - \delta)$

采用 Sigmoid 函数作为核函数时，支持向量机实现的就是一种多层感知器神经网络，应用 SVM 方法，隐含层节点数目(它确定神经网络的结构)、隐含层节点对输入节点的权值都是在设计(训练)的过程中自动确定的。而且支持向量机的理论基础决定了它最终求得的是全局最优值而不是局部最小值，也保证了它对于未知样本的良好泛化能力而不会出现过学习现象。

在选取核函数解决实际问题时，通常采用的方法有：

一是利用专家的先验知识预先选定核函数；

二是采用 Cross-Validation 方法，即在进行核函数选取时，分别试用不同的核函数，归纳误差最小的核函数就是最好的核函数。如针对傅立叶核、RBF 核，结合信号处理问题中的函数回归问题，通过仿真实验，对比分析了在相同数据条件下，采用傅立叶核的 SVM 要比采用 RBF 核的 SVM 误差小很多。

三是采用由 Smits 等人提出的混合核函数方法，该方法较之前两者是目前选取核函数的主流方法，也

是关于如何构造核函数的又一开创性的工作。将不同的核函数结合起来后会有更好的特性，这是混合核函数方法的基本思想。

8. 带核的 SVM 为什么能分类非线性问题？

解析：

核函数的本质是两个函数的内积，而这个函数在 SVM 中可以表示成对于输入值的高维映射。注意核并不是直接对应映射，核只不过是一个内积。

9. 如果 SVM 模型欠拟合，以下方法哪些可以改进模型：（ ）

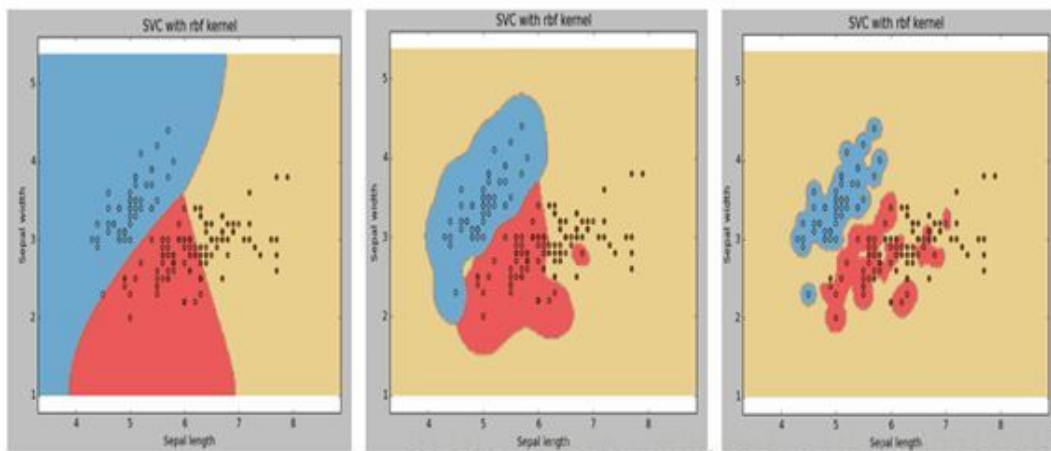
- A. 增大惩罚参数 C 的值
- B. 减小惩罚参数 C 的值
- C. 减小核系数(gamma 参数)

答案：（A）

解析：

如果 SVM 模型欠拟合，我们可以调高参数 C 的值，使得模型复杂度上升。

10. 下图是同一个 SVM 模型，但是使用了不同的径向基核函数的 gamma 参数，依次是 g1, g2, g3，下面大小比较正确的是：（ ）



- A. $g1 > g2 > g3$
- B. $g1 = g2 = g3$
- C. $g1 < g2 < g3$
- D. $g1 \geq g2 \geq g3$
- E. $g1 \leq g2 \leq g3$

答案：（C）

11. 假如我们使用非线性可分的 SVM 目标函数作为最优化对象，我们怎么保证模型线性可分？（ ）

- A. 设 $C=1$
- B. 设 $C=0$

- C. 设 C =无穷大
D. 以上都不对

答案: (C)

解析:

C 无穷大保证了所有的线性不可分都是可以忍受的。

12. 训练完 SVM 模型后, 不是支持向量的那些样本我们可以丢掉, 也可以继续分类: (A)

- A. 正确
B. 错误

答案: (A)

解析:

SVM 模型中, 真正影响决策边界的是支持向量。

13. 简述 LR 和 SVM 的联系与区别?

解析:

- 1、 LR 和 SVM 都可以处理分类问题, 且一般都用于处理线性二分类问题 (在改进的情况下可以处理多分类问题)
- 2、 两个方法都可以增加不同的正则化项, 如 $L1$ 、 $L2$ 等等。所以在很多实验中, 两种算法的结果是很接近的。

区别:

- 1、 LR 是参数模型, SVM 是非参数模型。
- 2、 从目标函数来看, 区别在于逻辑回归采用的是 Logistical Loss, SVM 采用的是 hinge loss. 这两个损失函数的目的都是增加对分类影响较大的数据点的权重, 减少与分类关系较小的数据点的权重。
- 3、 SVM 的处理方法是只考虑 Support Vectors, 也就是和分类最相关的少数点, 去学习分类器。而逻辑回归通过非线性映射, 大大减小了离分类平面较远的点的权重, 相对提升了与分类最相关的数据点的权重。
- 4、 逻辑回归相对来说模型更简单, 好理解, 特别是大规模线性分类时比较方便。而 SVM 的理解和优化相对来说复杂一些, SVM 转化为对偶问题后, 分类只需要计算与少数几个支持向量的距离, 这个在进行复杂核函数计算时优势很明显, 能够大大简化模型和计算。
- 5、 Logic 能做的 SVM 能做, 但可能在准确率上有问题, SVM 能做的 Logic 有的做不了。

14. 简述 $L1$ 和 $L2$ 的区别。

解析:

$L1$ 范数 ($L1$ norm) 是指向量中各个元素绝对值之和, 也有个美称叫“稀疏规则算子” (Lasso regularization)。

比如 向量 $A=[1, -1, 3]$, 那么 A 的 $L1$ 范数为 $|1|+|-1|+|3|$ 。

简单总结一下就是:

$L1$ 范数: 为 x 向量各个元素绝对值之和。

L2 范数: 为 x 向量各个元素平方和的 $1/2$ 次方, L2 范数又称 Euclidean 范数或 Frobenius 范数

L_p 范数: 为 x 向量各个元素绝对值 p 次方和的 $1/p$ 次方.

在支持向量机学习过程中, L1 范数实际是一种对于成本函数求解最优的过程, 因此, L1 范数正则化通过向成本函数中添加 L1 范数, 使得学习得到的结果满足稀疏化, 从而方便人类提取特征。

L1 范数可以使权值稀疏, 方便特征提取。

L2 范数可以防止过拟合, 提升模型的泛化能力。

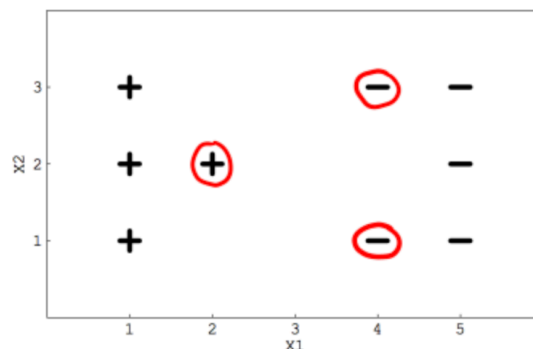
面试中会遇到的, L1 和 L2 正则先验分别服从什么分布, L1 是拉普拉斯分布, L2 是高斯分布。

15. 现有一个点能被正确分类且远离决策边界。如果将该点加入到训练集, 为什么 SVM 的决策边界不受影响, 而已经学好的 logistic 回归会受影响?

解析:

因为 SVM 采用的是 hinge loss, 当样本点被正确分类且远离决策边界时, SVM 给该样本的权重为 0, 所以加入该样本决策边界不受影响。而 logistic 回归采用的是 log 损失, 还是会给该样本一个小小的权重。

16. 假设有一个线性 SVM 分类器用来处理二分类问题, 下图显示给定的数据集, 其中被红色圈出来的代表支持向量。



- 1) 若移动其中任意一个红色圈出的点, 决策边界是否会变化? ()
A. 会
B. 不会

答案: A

解析:

这三个支持向量确定决策边界的位置, 如果它们中的任意一个被改变, 决策边界一定会变化。

- 2) 若移动其中任意一个没有被圈出的点, 决策边界会发生变化? ()
A. 会
B. 不会

答案: B

解析:

支持向量以外的点并不会影响决策边界。

17. SVM 中的泛化误差代表什么？（ ）

- A. 分类超平面与支持向量的距离
- B. SVM 对新数据的预测准确度
- C. SVM 中的误差阈值

答案：B

解析：

泛化误差在统计学上的意义是“样本外误差”。这是一种模型在未知的新数据上预测准确性的度量。

18. 若参数 C (cost parameter) 被设为无穷，下面哪种说法是正确的？（ ）

- A. 只要最佳分类超平面存在，它就能将所有数据全部正确分类
- B. 软间隔 SVM 分类器将正确分类数据
- C. 二者都不对

答案：A

解析：

在如此高的误分类惩罚下，不会存在软间隔分类超平面，因为一点错误都不可能发生。

19. 怎样理解“硬间隔”？（ ）

- A. SVM 只允许极小误差
- B. SVM 允许分类时出现一定范围的误差
- C. 二者都不对

答案：A

解析：

硬间隔表明 SVM 对分类正确性的要求非常严格，所以模型会尽力在训练集上表现的更好，这通常会造成过拟合。

20. SVM 算法的最小时间复杂度是 $O(n^2)$ ，基于此，以下哪种规格的数据集并不适该算法？（ ）

- A. 大数据集
- B. 小数据集
- C. 中等数据集
- D. 不受数据集大小影响

答案：A

解析：

除了规模要小，具有明显分类边界的数据集也更适合 SVM 算法。

21. SVM 算法的性能取决于：

- A. 核函数的选择
- B. 核函数的参数

- C. 软间隔参数 C
- D. 以上所有

答案: D

解析:

上述三点都会影响到算法的表现, 应尽量选择最佳的参数, 以最大限度提高效率、减少误差以及避免过拟合。

22. 支持向量是最靠近决策表面的数据点

- A. 正确
- B. 错误

答案: A

解析:

支持向量是最接近超平面的点, 这些点也最难分类, 他们会直接影响决策边界的位置。

23. 以下哪种情况会导致 SVM 算法性能下降?

- A. 数据线性可分
- B. 数据干净、格式整齐
- C. 数据有噪声, 有重复值

答案: C

解析:

当数据集有大量噪声和重叠点时, 要想得到一个清晰的分类超平面非常困难。

24. 假设你选取了高 Gamma 值的径向基核 (RBF), 这表示:

- A. 建模时, 模型会考虑到离超平面更远的点
- B. 建模时, 模型只考虑离超平面近的点
- C. 模型不会被数据点与超平面的距离影响

答案: B

解析:

Gamma 参数会调整远离超平面的数据点对模型的影响。

Gamma 值较低, 模型受到很多约束, 会包含训练集中所有数据点, 并不会捕捉到真正的模式。

Gamma 值较高, 模型对数据集形状的勾勒更加有效。

25. SVM 中的代价参数 C 表示什么?

- A. 交叉验证的次数
- B. 用到的核函数
- C. 在分类准确性和模型复杂度之间的权衡
- D. 以上都不对

答案: C

解析：

代价参数的大小决定了 SVM 能允许的误分类程度。

C 的值小：优化的目标是得到一个尽可能光滑的决策平面。

C 的值大：模型只允许出现很小数量的误分类点。

它可以简单的看做是对误分类的惩罚。

26. 假定有一个数据集 S，但该数据集有很多误差（这意味着不能太过依赖任何特定的数据点）。若要建立一个 SVM 模型，它的核函数是二次多项式核，同时，该函数使用变量 C（cost parameter）作为一个参数。

1) 若 C 趋于无穷，以下哪种说法正确？

- A. 数据仍可正确分类
- B. 数据无法正确分类
- C. 不确定
- D. 以上都不对

答案：A

解析：

若变量 C 的值很大，说明误分类的惩罚项非常大，优化的目标应该是让分类超平面尽量将所有的数据点都正确分类。

2) 若 C 的值很小，以下哪种说法正确？

- A. 会发生误分类现象
- B. 数据将被正确分类
- C. 不确定
- D. 以上都不对

答案：A

解析：

因为误分类的惩罚项非常小，模型得出的分类面会尽可能将大多数数据点正确分类，但有部分点会出现误分类现象。

27. 若训练时使用了数据集的全部特征，模型在训练集上的准确率为 100%，验证集上准确率为 70%。出现的问题是？

- A. 欠拟合
- B. 过拟合
- C. 模型很完美

答案：B

解析：

在训练集上准确率高，但在测试集上表现差是典型的过拟合现象。

28. 下面哪个是 SVM 在实际生活中的应用？

- A. 文本分类

- B. 图片分类
- C. 新闻聚类
- D. 以上都对

答案: D

解析:

SVM 在实际生活中的应用领域非常广泛, 从分类、聚类到手写字体识别都有涉及。

29. 假定你现在训练了一个线性 SVM 并推断出这个模型出现了欠拟合现象。

1) 在下一训练时, 应该采取下列什么措施?

- A. 增加数据点
- B. 减少数据点
- C. 增加特征
- D. 减少特征

答案: C

解析:

最好的选择就是生成更多的特征。

2) 假定你上一道题回答正确, 那么根本上发生的是:

- 1 偏差 (bias) 降低
 - 2 方差 (variance) 降低
 - 3 偏差增加
 - 4 方差增加
- A. 1 和 2
 - B. 2 和 3
 - C. 1 和 4
 - D. 2 和 4

答案: C

解析:

如果要使得模型性能更好, 就要在偏差和方差之间做出权衡。

泛化误差 = 偏差 + 方差 + 误差

误差: 由数据本身或模型本身的问题引起, 是期望泛化误差的下界。

偏差: 描述预测值的期望和真实值之间的差距, 度量模型本身拟合能力。

方差: 描述预测值的变化范围, 离散程度。度量了同样大小的训练集的变动导致的学习性能的变化。

度量模型对学习样本的依赖性。

一般模型越复杂, 学习能力越强, 误差会越小但方差越大。反之模型越简单, 对数据的拟合能力越弱, 误差大同时方差小。

3) 还是上面的问题, 如果不在特征上做文章, 而是改变一个模型的参数, 使得模型效果改善, 以下哪种方法是正确的?

- A. 增加代价参数 C

- B. 减小代价参数 C
- C. 改变 C 的值没有作用
- D. 以上都不对

答案: A

解析:

增加参数 C 的值会确保得到正则化的模型。

30. 在应用高斯核 SVM 之前, 通常都会对数据做正态化 (normalization), 下面对特征正态化的说法哪个是正确的?

- 1 对特征做正态化处理后, 新的特征将主导输出结果
- 2 正态化不适用于类别特征
- 3 对于高斯核 SVM, 正态化总是有用

- A. 1
- B. 1 和 2
- C. 1 和 3
- D. 2 和 3

答案: B

31. 假定现在有一个四分类问题, 你要用 One-vs-all 策略训练一个 SVM 的模型。请看下面的问题:

1) 由题设可知, 你需要训练几个 SVM 模型?

- A. 1
- B. 2
- C. 3
- D. 4

答案: D

解析:

多分类问题中, One-vs-all 策略要求为每一个类建立唯一的分类器, 属于此类的所有样例均为正例, 其余全部为负例。

2) 假定数据集中每一类的分布相同, 且训练一次 SVM 模型需要 10 秒, 若完成上面的任务, 共花费多少秒?

- A. 20
- B. 40
- C. 60
- D. 80

答案: B

解析:

花费时间为 $10 \times 4 = 40$ 秒。

3) 现在问题变了, 如果目前只需要将数据集分为 2 类, 需要训练多少次?

- A. 1
- B. 2
- C. 3
- D. 4

答案: A

解析:

该情况下训练 SVM 一次就能获得满意的结果。

32. 假定你使用阶数为 2 的线性核 SVM, 将模型应用到实际数据集上后, 其训练准确率和测试准确率均为 100%。

1) 假定现在增加模型复杂度 (增加核函数的阶), 会发生以下哪种情况?

- A. 过拟合
- B. 欠拟合
- C. 什么都不会发生, 因为模型准确率已经到达极限
- D. 以上都不对

答案: A

解析:

增加模型的复杂度会导致过拟合现象, 这与模型当前的状态无关。

2) 在增加了模型复杂度之后, 你发现训练准确率仍是 100%, 原因可能是? 、

- 1. 数据是固定的, 但我们在不断拟合更多的多项式或参数, 这会导致算法开始记忆数据中的所有内容
- 2. 由于数据是固定的, SVM 不需要在很大的假设空间中搜索

- A. 1
- B. 2
- C. 1 和 2
- D. 二者都不对

答案: C

33. 下面关于 SVM 中核函数的说法正确的是?

- 1 核函数将低维空间中的数据映射到高维空间
- 2 它是一个相似度函数

- A. 1
- B. 2
- C. 1 和 2
- D. 以上都不对

答案: C

34. SVM 的原理是什么？

解析：

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。（间隔最大是它有别于感知机）

（1）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；

（2）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；

（3）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。

注：以上各 SVM 的数学推导应该熟悉：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）。

35. SVM 为什么采用间隔最大化？

解析：

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。

然后应该借此阐述，几何间隔，函数间隔，及从函数间隔—>求解最小化 $\frac{1}{2}\|w\|^2$ 时的 w 和 b 。即线性可分支持向量机学习算法—最大间隔法的由来。

36. 为什么要将求解 SVM 的原始问题转换为其对偶问题？

解析：

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、自然引入核函数，进而推广到非线性分类问题。

37. 为什么 SVM 对缺失数据敏感？

解析：

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（决策树有）。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。

38. svm RBF 核函数的具体公式？

解析：

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right)$$

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。

39. 为什么 SVM 要引入核函数？

解析：

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。

引入映射后的对偶问题：

$$\begin{aligned} \max_{\lambda} \quad & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \phi^T(x_i) \phi(x_j) + \sum_{i=1}^N \lambda_i \\ \text{s.t.} \quad & \sum_{i=1}^N \lambda_i y_i = 0 \\ & 0 \leq \lambda_i \leq C, \forall i=1, 2, \dots, N \end{aligned}$$

在学习预测中，只定义核函数 $K(x, y)$ ，而不是显式的定义映射函数 ϕ 。因为特征空间维数可能很高，甚至可能是无穷维，因此直接计算 $\phi(x) \cdot \phi(y)$ 是比较困难的。相反，直接计算 $K(x, y)$ 比较容易（即直接在原来的低维空间中进行计算，而不需要显式地写出映射后的结果）。

核函数的定义： $K(x, y) = \langle \phi(x), \phi(y) \rangle$ ，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。

除了 SVM 之外，任何将计算表示为数据点的内积的方法，都可以使用核方法进行非线性扩展。