

· 改变任意一个的位置都会引入依赖变量· 决策边界就会发生变化。

· 决策边界只会与支持向量有关, 与其它无关。

SVM的泛化误差是对未知数据的预测能力。

硬间隔意味SVM对分类很严格, 在训练集上表现可能好, 有可能会过拟合。

模式识别与机器学习随堂测试

1. 朴素贝叶斯是一个 (A, D)。

- (A) 低方差模型 (B) 高方差模型 (C) 判别式模型 (D) 生成式模型

2. 对决策树模型, 关于其超参数“树的最大深度”, 下面说法哪些正确? A, C

- (A) 如果验证准确率相同, 值越低越好; \Rightarrow 时空复杂度会降低
(B) 如果验证准确率相同, 值越高越好;
(C) 值增加可能会导致过拟合; 更加细化(过拟合)
(D) 值增加可能会导致欠拟合。

3. 在L1正则的线性回归模型中, 如果正则参数很大, 会发生什么? A

梯度下降法与最小二乘法可以L1正则线性回归预测。

- (A) 一些系数将变为零, 增强模型的泛化能力。 $J(\theta) = \frac{1}{2}(\mathbf{X}\theta - \mathbf{Y})^T(\mathbf{X}\theta - \mathbf{Y}) + \alpha\|\theta\|_1$ (L1正则)

- (B) 一些系数将接近零, 但不是绝对零

其中n为样本个数, α 为常数系数, 需要进行调优。 $\|\theta\|_1$ 为L1范数。

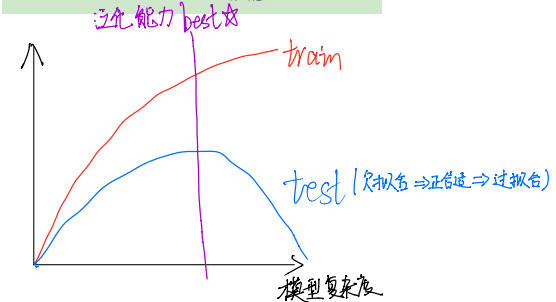
4. 阶数越高模型越复杂
对多项式回归而言, 哪一项对过拟合和欠拟合影响最大? A

- (A) 多项式的阶数
(B) 是否通过矩阵求逆/梯度下降学习权重
(C) 高斯噪声方差
(D) 每一次训练的输入个数固定

对于NN, 则是隐藏层输入节点数

$$J(\theta) = \frac{1}{2}(\mathbf{X}\theta - \mathbf{Y})^T(\mathbf{X}\theta - \mathbf{Y}) + \frac{1}{2}\alpha\|\theta\|_2^2 \quad (\text{L2正则})$$

其中 α 为常数系数, 需要进行调优。 $\|\theta\|_2$ 为L2范数。



5. 使用梯度下降训练Logistic回归分类器后, 您发现它对训练集欠拟合

, 在训练集或验证集上没有达到所需的性能。以下哪些项可能是有希望采取的步骤? C、D

- (A) 采用其他优化算法, 因为梯度下降得到的可能是局部最小值
(B) 减少训练样本
(C) 增加多项式特征值
(D) 改用较多隐含结点的神经网络模型

6. 在Logistic回归中, 关于一对其他 (One vs. Rest) 方法, 以下哪个选项是正确? B (A也可以)

- (A) 我们需要在类分类问题中拟合个模型
(B) 我们需要拟合个模型来分类类
(C) 我们只需要拟合1个模型来分类类
(D) 这些都不是

7. SVM的有效性取决于: D

径向基函数: RBF

γ : 定义单个训练样本的影响大小, 值越小影响越大
可视为样本影响半径的倒数.

C: 在误差与分类样本和分界面间做权衡
高C值, 不放弃任何点, \Rightarrow 增大C会得到正则化模型

- (A) 核函数选择
- (B) 核函数的参数
- (C) 软边距参数
- (D) 以上所有

γ : 衡量距离超平面远近点的影响

• SVM的代价参数表示:
误差与模型复杂度间的平衡.

• 更好的模型: 降低偏差, 提高方差

8. 当出现下述哪种情况时, SVM性能不佳? C

- (A) 数据线性可分
- (B) 数据干净
- (C) ~~数据有噪声~~

RBF 高斯核函数

9. 假设您在SVM中使用具有高(核函数宽度的倒数)值的RBF核函数。这意味着什么? C

- ~~(A)~~ 模型将考虑离超平面很远的点.
 - ~~(B)~~ 模型将只考虑离超平面很近的点.
 - (C) 模型不受点到超平面距离的影响。
 - (D) 以上都不对
- γ C值 高 γ 值, 影响小

10. 假设您采用一个线性SVM模型来处理某个任务, 并且知道这个SVM模型是欠拟合的。

下列哪些方法可以提升该模型性能? C、E

- (A) 减少训练样本
- (B) ~~增加训练样本~~
- (C) 增加特征
- (D) ~~减少特征~~
- (E) 增加参数
- (F) ~~减少参数~~

11. 如果我使用数据集的所有特征, 在训练集的准确率达到100%, 而验证集的准确率为

70%, 那么我应该注意什么? C

- (A) 模型是欠拟合的;
- (B) 模型是完美的;
- (C) ~~模型是过拟合的。~~

过拟合

One-vs-all模型: 每次把一个类当作正类, 其它类当作负类