### 支持向量机1（SVM）

1. SVM是监督学习
2. SVM可以做回归也可以做分类
3. SVM被称为大间距分类器，因为它寻求的是’分类间隔最大化’，因此具有很好的鲁棒性(健壮性)，泛化错误低，泛化能力强，适能力
4. SVM的优化目标:

IMG_256

C = 1/λ C大λ小(过拟合)，C小λ大(欠拟合)

C表示惩罚措施，如果惩罚力度过大则不敢犯错，所以过拟合。换之则相反

当 y=1时，z=Xθ>=1, 函数cost1(z)函数值为0

当 y=0时，z=Xθ<= -1, 函数cost0(z)函数值为0

值越远代价越小

1. 惩罚措施

当C不是非常非常大的时候，他可以忽略掉一些异常点的影响(忽略不重要的元素点,让其泛化线最为正确,如果C过大,则不是最优线)，得到更好的决策界。甚至当你的数据不是线性可分的时候，支持向量机也可以给出好的结果(核函数)

C较大时，λ较小，可能导致过拟合，高方差

C较小时，λ过大，可能导致欠拟合，高偏差

### 二、SVM2(最大间距分类器)

1.核函数可以使用低维数据进行运算，但其结果映射高维。(增高维度，从而线性可分)

学习思路:

1. 介绍2.内部流程3.解决问题
2. 高斯核函数是径向基函数(rbf）中的一种，代表相似度，公式：

X(样本)与L(地标点)之间自己的相似度，σ表示影响度。

1. SVM使用核函数生成新特征，然后用新特征进行分类预测，其优化目标



1. 常用的核函数由:高斯核函数,多项式核函数，字符串核函数，卡方核函数，直方图交集核函数，sigmoid核函数
2. 两个参数C和σ的影响
3. C较大时，相当于λ较小，可能过拟合，低偏差，高方差
4. C较小时，相当于λ较大，可能欠拟合，高偏差，低方差
5. σ较大时，特征值f(l)变化非常稳定，影响幅度大而平缓，高偏差，低方差
6. σ较小时，特征值f(l)变化不稳定，影响幅度小而不平衡，低偏差，高方差
7. SVM普遍使用的准则:
8. 当m<n，m表示样本数，n表示特征数。即训练集数据量不够支持训练一个复杂的非线性模型。
9. 当m>n,且n较小，m中等。列如:n在1-1000之间，而m在10-10000之间，使用高斯核函数支持向量机。
10. 当m>n时，n较小，m较大，列如:n在1-1000之间，m大于50000，则使用支持向量机则非常慢，解决方法是创造、增加更多特征，然后使用逻辑回归，或者不带核函数的支持向量机
11. SVM算法的主要优点:
12. 解决高纬度的分类问题和回归问题很有效，在特征维度大于样本数时仍然有效
13. 仅仅使用一部分支持向量(决定最大间距的元素)来做超平面的决策，无需依赖全部数据。
14. 有大量的核函数可以使用，从而可以灵活解决各种非线性分类问题。
15. 样本量不是海量数据时，分类准确率过高，泛化能力强
16. SVM算法的主要缺点有:
17. 如果特征数原因大于样本数，则SVM表现一般
18. SVM的样本量非常大，核函数映射维度非常高时，计算量过大，不适合使用
19. 非线性问题的核函数选择没有通用标准。难以选择一个合适的核函数
20. SVM对缺失数据敏感