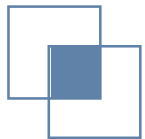




机器学习理论与工程实践

第4章 支持向量机

徐文星



目录



1 支持向量的凸优化描述

2 对偶问题

3 核函数

4 软间隔与正则化



SVM算法思想

2021-03-03 11:24:57 | 10 经验 | 0 人已查看



SVM数学描述

2021-03-03 11:41:05 | 10 经验 | 0 人已查看



SVM对偶问题 I

2021-03-03 11:56:39 | 10 经验 | 0 人已查看



SVM对偶问题 II

2021-03-03 11:57:14 | 10 经验 | 0 人已查看



SVM核函数

2021-03-03 11:50:47 | 10 经验 | 0 人已查看



SVM常用核函数

2021-03-03 12:02:07 | 10 经验 | 0 人已查看



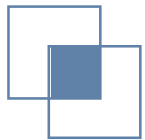
SVM软间隔

2021-03-03 11:47:56 | 10 经验 | 0 人已查看

<https://www.bilibili.com/video/BV1dJ411B7gh?p=6>

~

<https://www.bilibili.com/video/BV1dJ411B7gh?p=12>

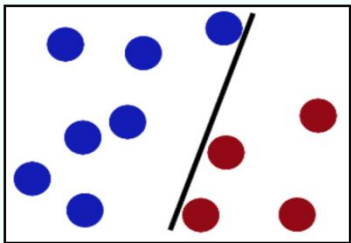


SVM (支持向量机)

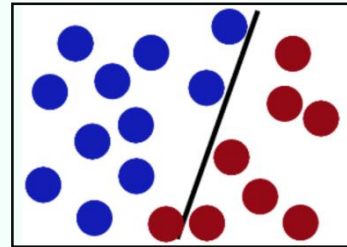


在很久以前的情人节，大侠要去救他的爱人，但魔鬼和他玩了一个游戏。

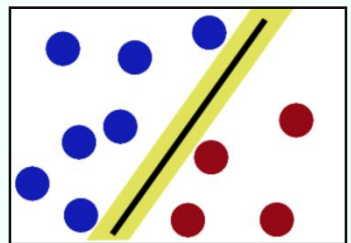
1. 魔鬼在桌子上似乎有规律放了两种颜色的球，说：“你用一根棍分开它们？要求：尽量在放更多球之后，仍然适用。”于是大侠这样放，干的似乎还不错？



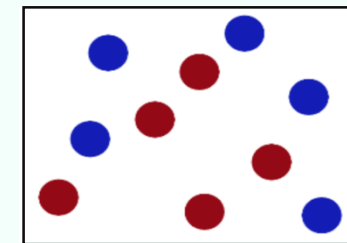
2. 然后魔鬼，又在桌上放了更多的球，似乎有一个球站错了阵营。



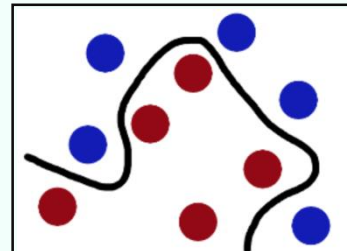
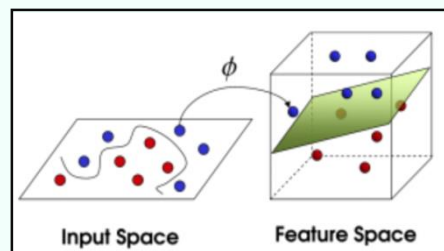
3. 大侠想到办法了，就是试图把棍放在最佳位置，好让在棍的两边有尽可能大的间隙，取名SVM。



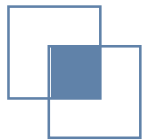
4. 魔鬼看到大侠已经学会了一个trick，于是魔鬼给了大侠一个新的挑战。



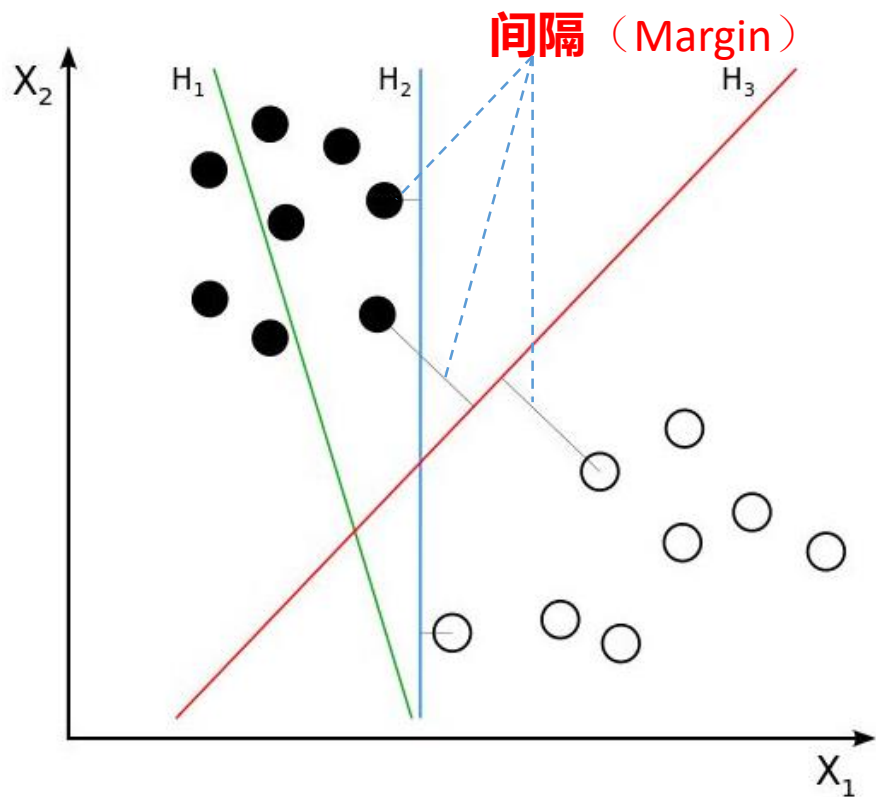
5. 现在，大侠没有棍可以很好帮他分开两种球了，现在怎么办呢？当然像所有武侠片中一样大侠一拍桌子，球飞到空中。然后，凭借大侠的轻功，抓起一张纸，插到了两种球的中间。



再之后，人们把这些球叫做 **「data」**，把棍子叫做 **「classifier」**，最大间隙trick 叫做 **「optimization」**，拍桌子叫做 **「kernelling」**，那张纸叫做 **「hyperplane」**。

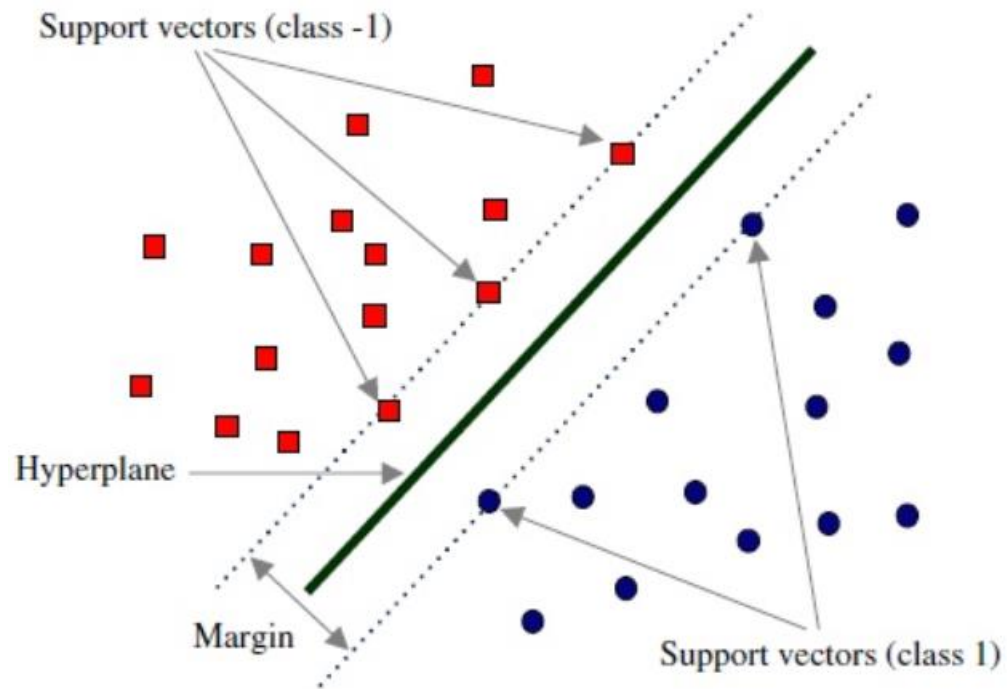


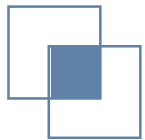
SVM (支持向量机)



直观理解： 对一个数据点进行分类，当分割线（超平面）离数据点的“间隔”越大，分类的确信度（confidence）也越大。

支持向量机： 中间直线就是分类超平面，不难发现，用来确定这条直线其实只需要两条虚线上的几个点就够了。





SVM (支持向量机)



支持向量机 (Support Vector Machine)
是一种有监督学习方法，它尝试寻找一个最优决策边界，使距离两个类别最近的样本最远。

硬间隔支持向量机：

$$\begin{aligned} & \max \frac{1}{\|w\|} \\ & \text{等价} \quad \min \|w\| \quad \min \frac{1}{2} \|w\|^2 \\ & \text{s.t.} \quad \begin{cases} \frac{w^T x^{(i)} + b}{\|w\|} \geq \frac{1}{\|w\|} & \forall y^{(i)} = 1 \\ \frac{w^T x^{(i)} + b}{\|w\|} \leq -\frac{1}{\|w\|} & \forall y^{(i)} = -1 \end{cases} \\ & \text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{aligned}$$

硬间隔（线性可分）支持向量机最优化问题：

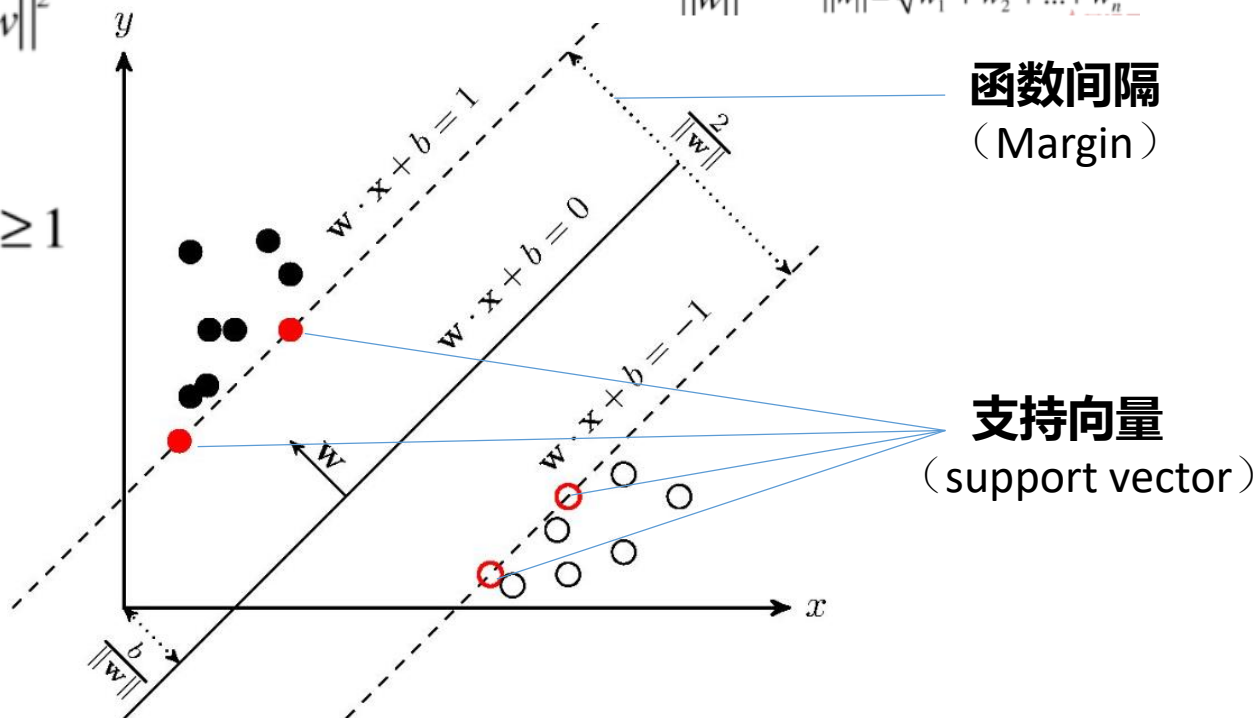
$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 \\ & \text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{aligned}$$

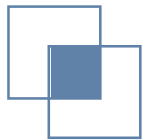
回忆解析几何，点到直线的距离

$$(x, y) \text{ 到 } Ax + By + C = 0 \text{ 的距离} \quad \frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$

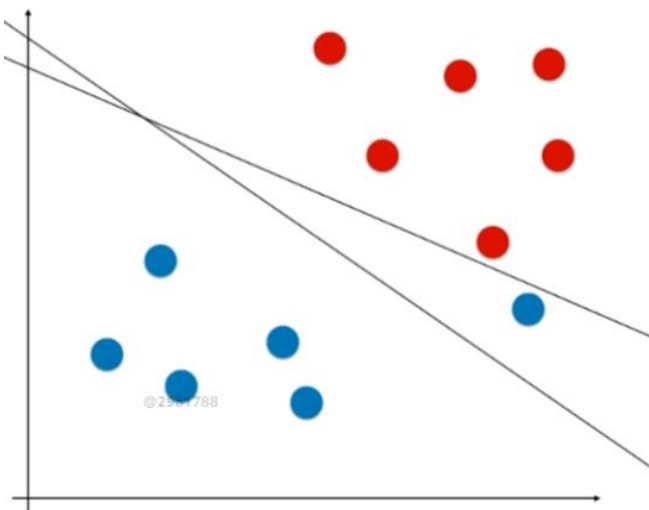
拓展到n维空间 $\theta^T x_b = 0 \rightarrow w^T x + b = 0$

$$\frac{|w^T x + b|}{\|w\|} \quad \|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$$



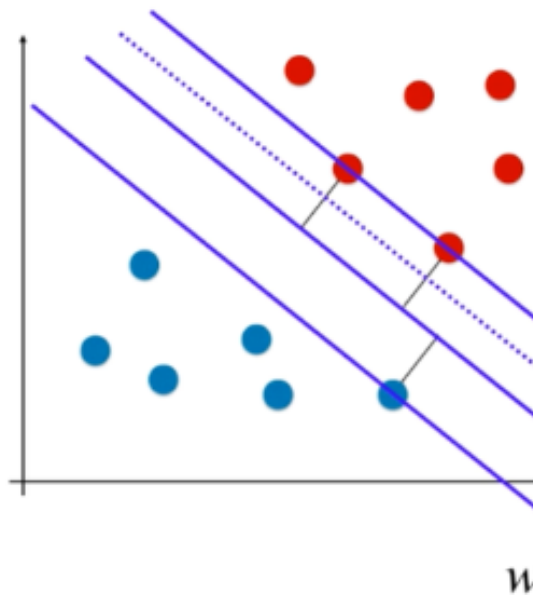


SVM (支持向量机)



SVM基本型
(硬间隔支持向量机) :

$$\min \frac{1}{2} \|w\|^2$$
$$s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1$$



软间隔支持向量机:

$$\min \frac{1}{2} \|w\|^2 + \boxed{C} \sum_{i=1}^m \zeta_i$$
$$s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i$$
$$\zeta_i \geq 0$$

L1正则

$$\min \frac{1}{2} \|w\|^2 + \boxed{C} \sum_{i=1}^m \zeta_i^2$$

$$s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i$$
$$\zeta_i \geq 0$$

L2正则

$$\min \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \zeta_i$$

$$s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i$$

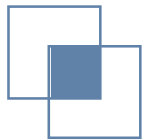
$$\zeta_i \geq 0$$

$$w^T x + b = 1$$

$$w^T x + b = 0$$

$$w^T x + b = -1$$

$$w^T x + b = 1 - \zeta$$



SVM (支持向量机)



SVM算法终极形态:

① 训练流程

输入 $\{(x_i, y_i)\}, i=1 \sim m$

求解优化问题:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & \alpha_i \geq 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

算b

$$b = \frac{1}{|S|} \sum_{s \in S} \left(y_s - \sum_{i \in S} \alpha_i y_i \kappa(x_i, x_s) \right)$$

$$S = \{i \mid \alpha_i > 0, i = 1, 2, \dots, m\}$$

可得模型

$$f(x) = \sum_{i=1}^m \alpha_i y_i \kappa(x, x_i) + b$$

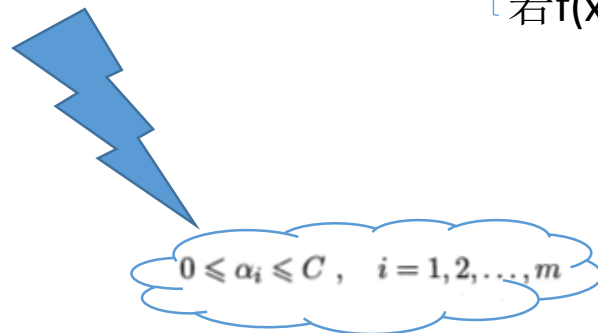
② 测试流程

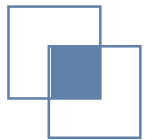
输入测试样本x

计算f(x)

$$f(x) = \sum_{i=1}^m \alpha_i y_i \kappa(x, x_i) + b$$

若 $f(x) \geq 0$, 则 $y = +1$
若 $f(x) < 0$, 则 $y = -1$



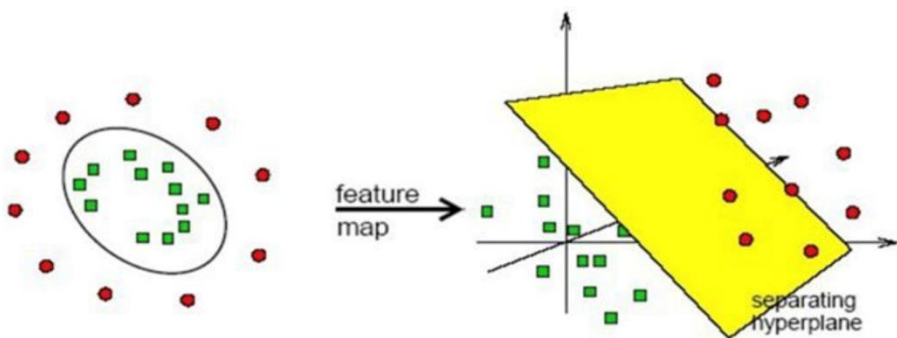


SVM (支持向量机)



支持向量机 (Support Vector Machine ,SVM) 是一种有监督学习方法，它尝试寻找一个最优决策边界，使距离两个类别最近的样本最远，从而对分类问题提供良好的泛化能力。

非线性支持向量机与核函数：



- 线性核函数： $K(x, z) = x \cdot z$
- 多项式核函数： $K(x, z) = (x \cdot z + 1)^p$
- 高斯核函数： $K(x, z) = \exp(-\frac{\|x-z\|^2}{2\sigma^2})$
- 混合核： $\lambda K_1(x, z) + (1 - \lambda)K_2(x, z), 0 \leq \lambda < 1$

SVM的优点：

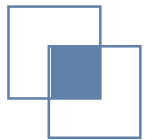
- ✓ 相对于其他训练分类算法**不需要过多样本**，并且由于SVM引入了核函数，所以SVM可以处理高维样本
- ✓ **结构风险最小**。这种风险是指分类器对问题真实模型的逼近与问题真实解之间的累积误差
- ✓ **非线性**，是指SVM擅长应付样本数据线性不可分的情况，主要通过松弛变量（也叫惩罚变量）和核函数技术来实现，这一部分也正是SVM的精髓所在。

常用软件工具包：

LibSVM: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

SVM-Light: <http://svmlight.joachims.org/>

Liblinear: <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

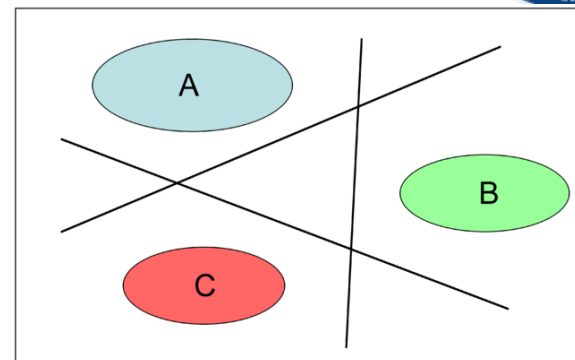


多分类SVM



一对多法 (one-versus-rest,简称OVR SVMs)

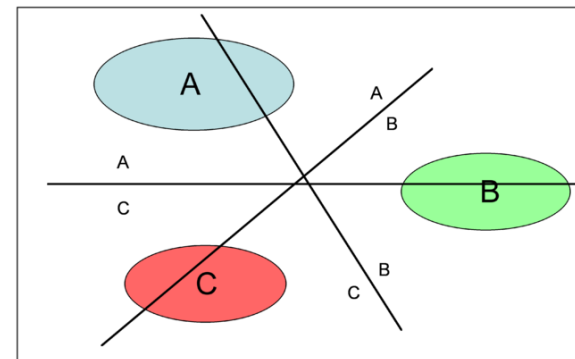
训练时依次把某个类别的样本归为一类,其他剩余的样本归为另一类,这样 k 个类别的样本就构造出了 k 个SVM。分类时将未知样本分类为具有最大分类函数值的那类。



一对一法 (one-versus-one,简称OVO SVMs或者pairwise)

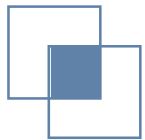
其做法是在任意两类样本之间设计一个SVM, 因此 k 个类别的样本就需要设计 $k(k-1)/2$ 个SVM。

当对一个未知样本进行分类时, 最后得票最多的类别即为该未知样本的类别。



层次支持向量机

层次分类法首先将所有类别分成两个子类, 再将子类进一步划分成两个次级子类, 如此循环, 直到得到一个单独的类别为止。



目录



1

支持向量机的应用



1-兵王问题

2021-02-19 16:08:17 | 10 经验 | 5 人已查看



2-LIBSVM解兵王问题

2021-02-19 16:08:17 | 10 经验 | 4 人已查看



3-兵王问题测试结果

2021-02-19 16:08:17 | 10 经验 | 4 人已查看



4-ROC曲线

2021-02-19 16:08:17 | 10 经验 | 4 人已查看

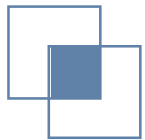


5-SVM处理多分类问题

2021-03-08 14:16:48 | 10 经验 | 0 人已查看

<https://www.bilibili.com/video/BV1dJ411B7gh?p=14> ~

<https://www.bilibili.com/video/BV1dJ411B7gh?p=18>



Tips



■ 混淆矩阵

		预测	
实际		正样本	负样本
	正样本	True Positive (TP)	False Negative (FN)
	负样本	False Positive (FP)	True Negative (TN)

TP: 将正样本识别为正样本的数量（或概率）

FN: 将正样本识别为负样本的数量（或概率）

FP: 将负样本识别为正样本的数量（或概率）

TN: 将负样本识别为负样本的数量（或概率）



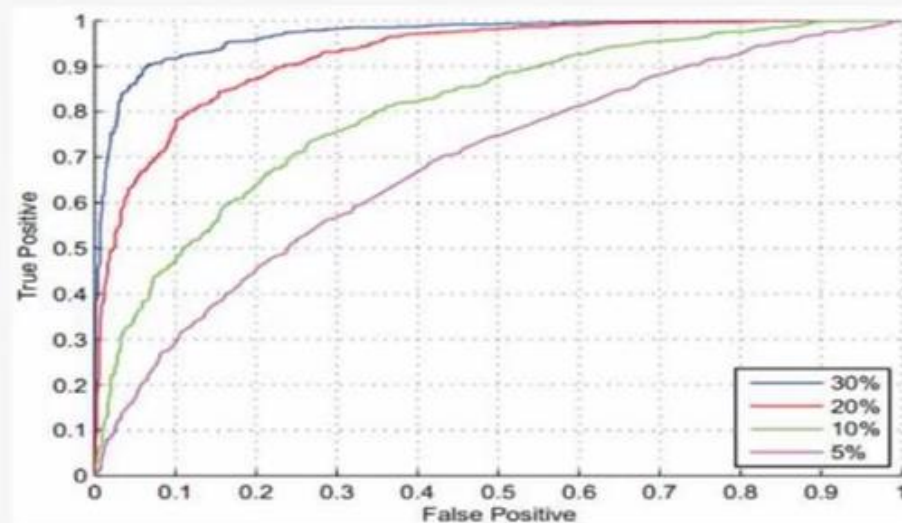
4-ROC曲线

2021-02-19 16:08:17 | 10 经验 | 4 人已查看

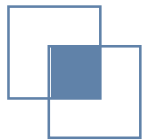


3-兵王问题测试结果

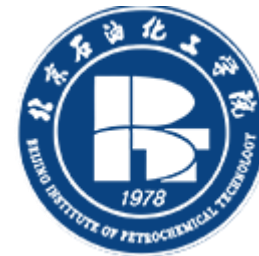
2021-02-19 16:08:17 | 10 经验 | 4 人已查看



ROC (Receiver Operating Character) 曲线，
是一条横坐标**FP**，纵坐标**TP**的曲线



感谢聆听



有什么问题吗？