

14 支持向量机

14 支持向量机

14.1 本集内容简介

14.2 支持向量机简介

14.3 线性分类器

14.4 分类间隔

14.5 线性可分的原问题

证明是凸优化

来看看slator条件

14.6 线性可分的对偶问题

14.7 线性可分的实验

14.9 线性不可分的对偶问题

14.10 KKT条件下的使用

14.11 线性不可分的实验

14.12 核函数实验

14.13 核映射与核函数

Mercer条件

14.14 本集总结

14.1 本集内容简介

支持向量机简介
线性分类器
分类间隔
线性可分问题
线性可分的对偶问题
线性不可分问题
线性不可分的对偶问题
核映射与核函数

14.2 支持向量机简介

支持向量机简介

由Vapnik等人提出，在出现后的20多年里它是最有影响力的机器学习算法之一

在深度学习技术出现之前，使用高斯核（RBF）的支持向量机在很多分类问题上一度取得了最好的结果

不仅可以用于分类问题，还可以用于回归问题

具有泛化性能好，适合小样本和高维特征等优点

支持向量机在机器学习里面，对数学要求比较高的。



- 对偶
- KKT

14.3 线性分类器

支持向量机是从线性分类器里衍生出来的。

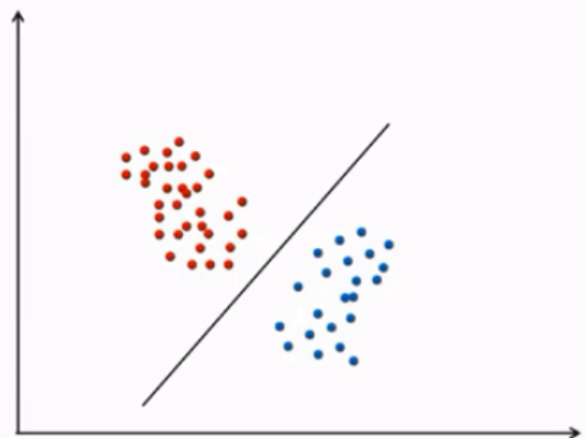
线性分类器

二分类问题，样本标签值为+1或-1

$$\mathbf{w}^T \mathbf{x} + b = 0$$

$$\text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

超平面的方程只是给出了分界面，哪边为正，哪边为负，是可以灵活控制的



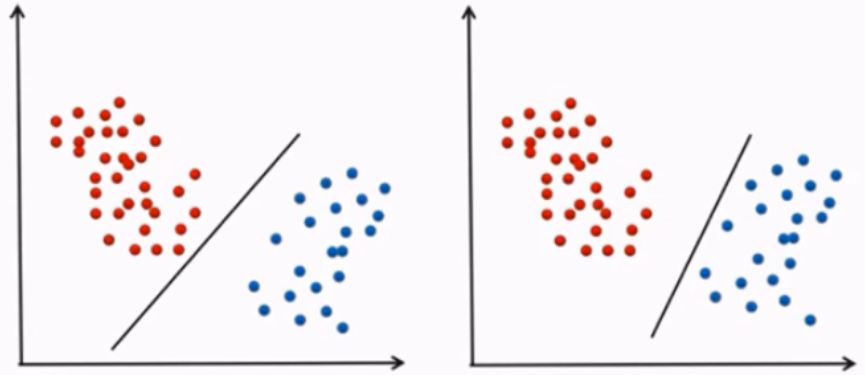
14.4 分类间隔

前面都是对二分类问题，如果对于多分类问题。

分类间隔

对于一个问题，可行的分类器不止一个，哪一个是最好的？

分类平面应该不偏向于任何一类，并且离两个类的样本都尽可能的远



尽可能使得分类间隔最大化的那个最好。

14.5 线性可分的原问题

线性可分的问题

首先要保证样本都被正确分类

对于正样本 $\mathbf{w}^T \mathbf{x} + b \geq 0$
对于负样本 $\mathbf{w}^T \mathbf{x} + b < 0$

可以统一写成 $y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 0$

其次是离两类样本尽可能远 $d = \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|}$

$$\min_i |\mathbf{w}^T \mathbf{x}_i + b| = 1$$

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1$$

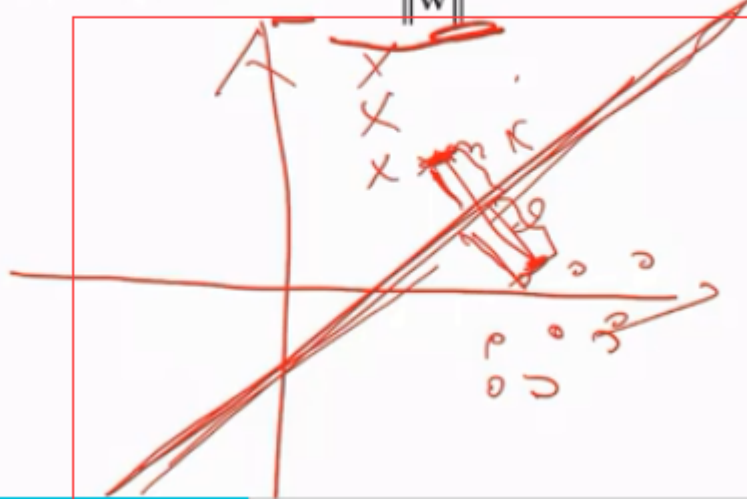
$$\begin{aligned} d(\mathbf{w}, b) &= \min_{\mathbf{x}_i, y_i = -1} d(\mathbf{w}, b; \mathbf{x}_i) + \min_{\mathbf{x}_i, y_i = 1} d(\mathbf{w}, b; \mathbf{x}_i) \\ &= \min_{\mathbf{x}_i, y_i = -1} \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|} + \min_{\mathbf{x}_i, y_i = 1} \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|} \\ &= \frac{1}{\|\mathbf{w}\|} \left(\min_{\mathbf{x}_i, y_i = -1} |\mathbf{w}^T \mathbf{x}_i + b| + \min_{\mathbf{x}_i, y_i = 1} |\mathbf{w}^T \mathbf{x}_i + b| \right) \\ &= \frac{2}{\|\mathbf{w}\|} \end{aligned}$$

离两类样本都远直线如下

两类样本距离尽可能的最大化

其次是离两类
样本尽可能远

$$d = \frac{|w^T x_i + b|}{\|w\|}$$



回忆一下解析几何里面，点到超平面的距离。

可以统一写成

$$y_i (w^T x_i + b) \geq 0$$

其次是离两类
样本尽可能远

$$d = \frac{|w^T x_i + b|}{\|w\|}$$

张样本的距离和负样本

$$\min_i |w^T x_i + b| = 1$$

$$y_i (w^T x_i + b) \geq 1$$

$$w^T x_i + b \geq 1$$

$$w^T x_i + b \leq -1$$

计算距离

$$\begin{aligned}
 d(w, b) &= \min_{x_i, y_i = -1} d(w, b; x_i) + \min_{x_i, y_i = 1} d(w, b; x_i) \\
 &= \min_{x_i, y_i = -1} \frac{|w^T x_i + b|}{\|w\|} + \min_{x_i, y_i = 1} \frac{|w^T x_i + b|}{\|w\|} \\
 &= \frac{1}{\|w\|} \left(\min_{x_i, y_i = -1} |w^T x_i + b| + \min_{x_i, y_i = 1} |w^T x_i + b| \right) \\
 &= \frac{2}{\|w\|}
 \end{aligned}$$

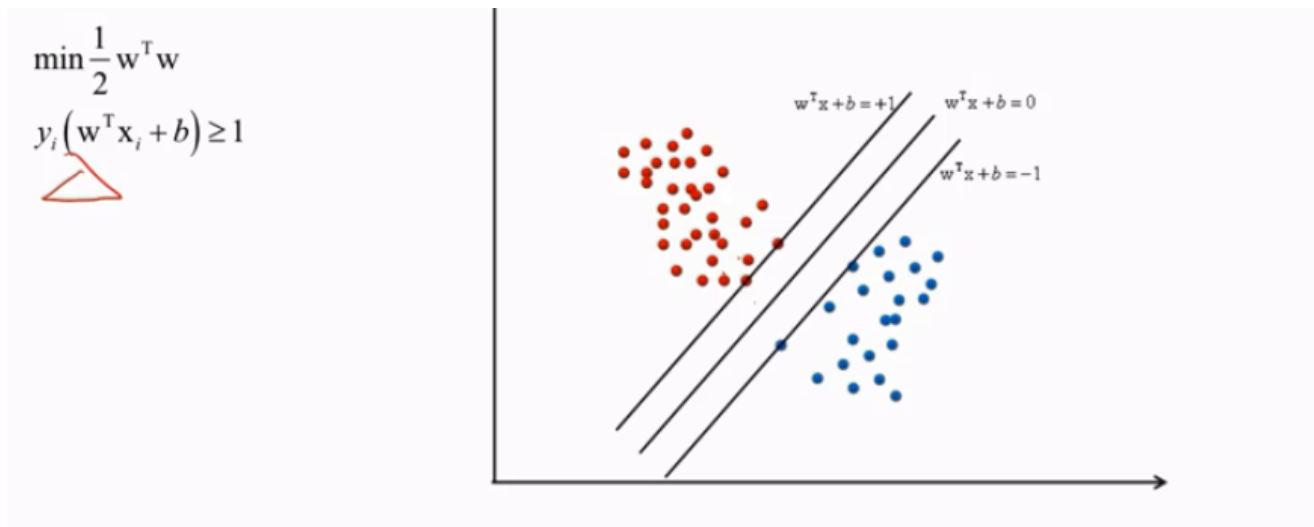
所以有，

通过范数优化

$$\begin{aligned}
 &\min \frac{1}{2} w^T w \\
 &y_i (w^T x_i + b) \geq 1 \\
 &\max \frac{2}{\|w\|} \\
 &\geq 0
 \end{aligned}$$

(Handwritten red circle around the minimization problem and the constraint, with arrows pointing to the handwritten notes above and below.)

看看示意图，红色为正样本

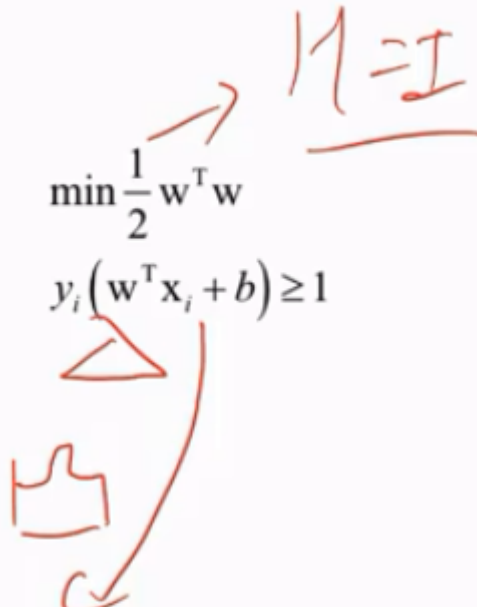


证明是凸优化


假设有一个优化目标

$$\min \frac{1}{2} w^T w$$

$$y_i (w^T x_i + b) \geq 1$$



$$g(x) \leq 0$$



线性可分围成的都是凸集

我们计算一下优化目标函数

$$\frac{1}{2}(w_1^2 + w_2^2 + \dots + w_n^2) \frac{\partial L}{\partial w_i} = 0 \quad \text{if } i \neq j$$

矩阵的形状如如，可以知道，这是一个严格凸集

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \text{严格凸}$$

来看看slator条件

1. 是凸优化
2. 至少存在图中公式2

Slator

1. 1.3

2. $g_i(K) < 0$

$g_d(X) < 0$

$\rightarrow \frac{\partial L}{\partial w_i} = 0$

如果原问题难求解，满足凸优化的条件，可以转化为对偶问题来求解。

14.6 线性可分的对偶问题

前面的最优化问题带有太多的不等式约束，不易求解

这个优化问题是凸优化，而且满足Slater条件，因此可以强对偶成立，可以用拉格朗日对偶转化成对偶问题

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i (w^T x_i + b) - 1)$$

$$\min_{w, b} \max_{\alpha} L(w, b, \alpha) \Leftrightarrow \max_{\alpha} \min_{w, b} L(w, b, \alpha)$$

$$\frac{\partial L}{\partial b} = 0 \quad \sum_{i=1}^l \alpha_i y_i = 0$$

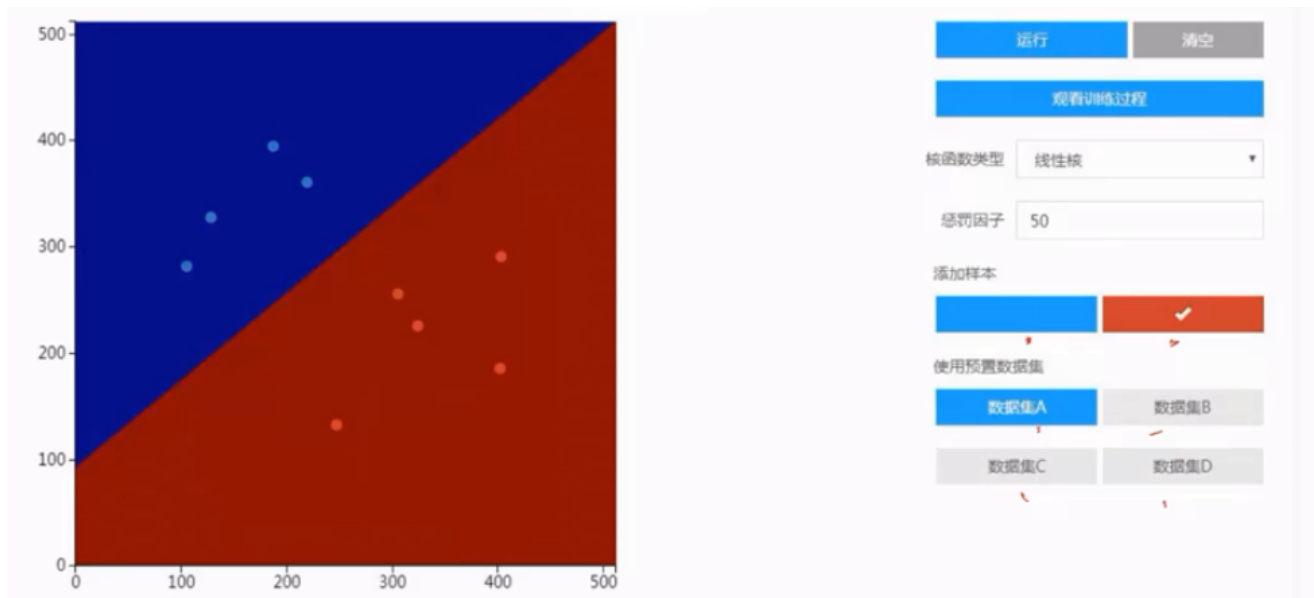
$$\nabla_w L = 0 \quad w = \sum_{i=1}^l \alpha_i y_i x_i$$

$$\max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{k=1}^l \alpha_k$$

回忆拉格朗日对偶，要到前面的章节去看。

14.7 线性可分的实验

直观看出把样本分开的是一条直线



14.9 线性不可分的对偶问题

转换成对偶问题

$$y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) \geq 1 - \xi_i \Rightarrow - \left(y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 + \xi_i \right) \leq 0$$

$$\xi_i \geq 0 \Rightarrow -\xi_i \leq 0$$

$$L(\mathbf{w}, b, \alpha, \xi, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i \left(y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 + \xi_i \right) - \sum_{j=1}^l \beta_j \xi_j$$

$$\frac{\partial L}{\partial b} = 0 \qquad \sum_{i=1}^l \alpha_i y_i = 0$$

$$\nabla_{\xi} L = 0 \qquad \alpha_i + \beta_i = C$$

$$\nabla_{\mathbf{w}} L = 0 \qquad \mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i$$

$$\max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{k=1}^l \alpha_k$$

$$\alpha_i + \beta_i = C \qquad \beta_i \geq 0 \qquad \alpha_i \leq C$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

$$0 \leq \alpha_i \leq C$$

$$y^T \alpha = 0$$

这是一个凸优化问题

$$Q = X^T X$$

$$X = [y_1 x_1, \dots, y_l x_l]$$

$$x^T Q x = x^T (X^T X) x = (X x)^T (X x) \geq 0$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

$$0 \leq \alpha_i \leq C$$

$$y^T \alpha = 0$$

这是一个凸函数，

Handwritten derivation showing the quadratic form $x^T Q x$ as a squared norm:

$$x^T Q x = x^T (X^T X) x = (X x)^T (X x)$$

The handwritten notes show the matrix X as a row vector of terms $[y_1 x_1, \dots, y_l x_l]$ and the vector x as a column vector. The final expression $(Xx)^T (Xx)$ is shown as a sum of squares, confirming it is non-negative.

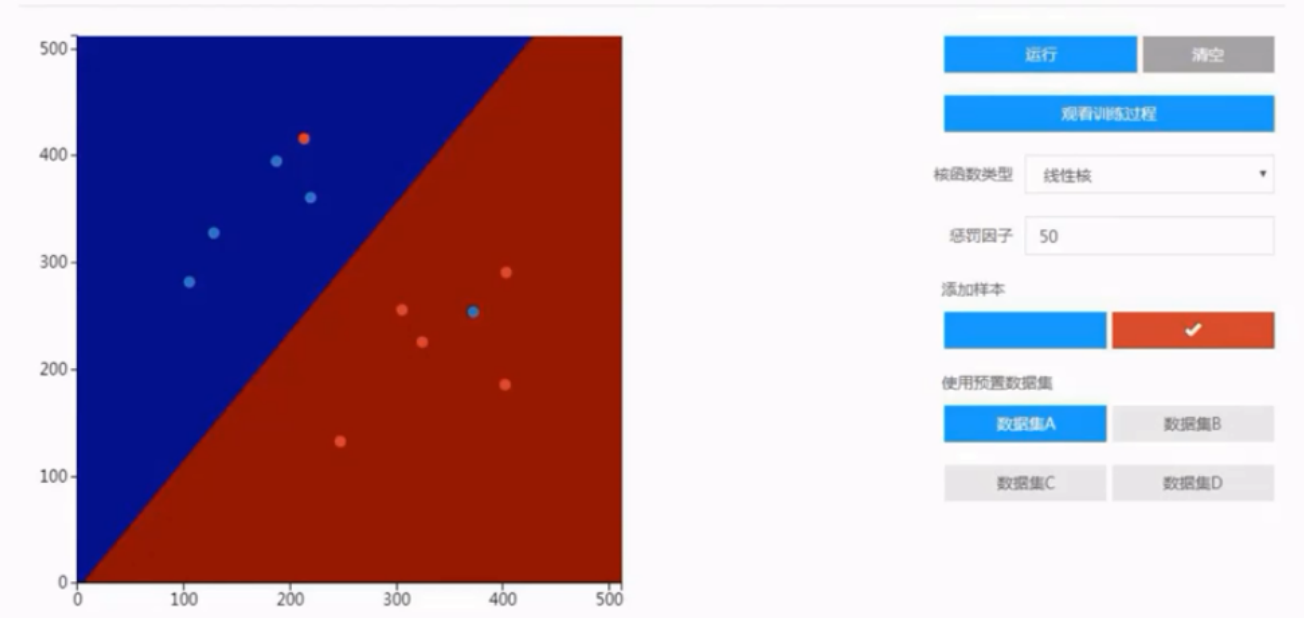
14.10 KKT条件下的使用

什么上界

自由向量

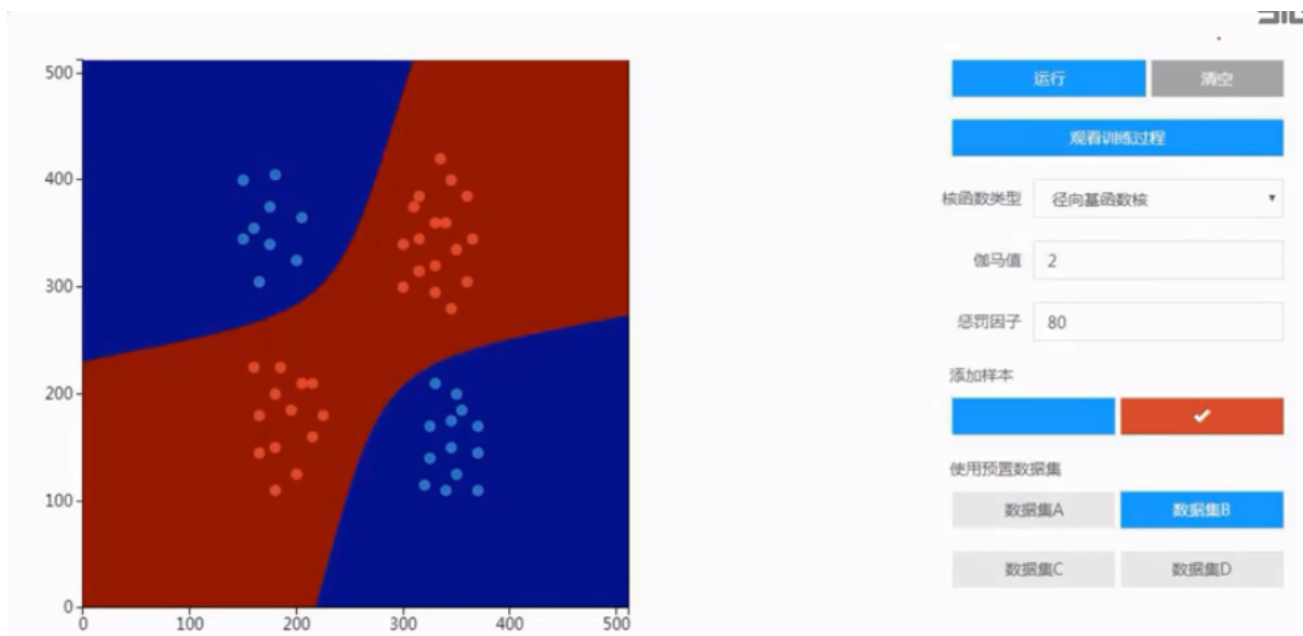
不是说的很抽象，是听得很不懂。不懂每个符号是怎么推导的。

14.11 线性不可分的实验



14.12 核函数实验

高斯核 (RBF)



14.13 核映射与核函数

前面通过松弛变量和惩罚因子，将线性

核映射与核函数

虽然引入了松弛变量和惩罚因子，可以处理线性不可分的问题，但SVM还是一个线性模型，只是允许错分样本的存在

核映射 $z = \phi(x)$

$$z_i^T z_j = \phi(x_i)^T \phi(x_j)$$

核函数 $K(x_i, x_j) = K(x_i^T x_j) = \phi(x_i)^T \phi(x_j)$

$$\text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i^T x) + b\right)$$

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) - \sum_{k=1}^l \alpha_k$$

$$0 \leq \alpha_i \leq C$$

$$\sum_{j=1}^l \alpha_j y_j = 0$$

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i^T x_j) - \sum_{k=1}^l \alpha_k$$

$$0 \leq \alpha_i \leq C$$

$$\sum_{j=1}^l \alpha_j y_j = 0$$

对偶问题仍然是凸优化问题

- 核映射
 - 先映射再做内积
- 核函数
 - 先做内积再做映射

核映射与核函数

虽然引入了松弛变量和惩罚因子，可以处理线性不可分的问题，但SVM还是一个线性模型，只是允许错分样本的存在

核映射 $z = \phi(x)$

$$z_i^T z_j = \phi(x_i)^T \phi(x_j)$$

核函数 $K(x_i, x_j) = K(x_i^T x_j) = \phi(x_i)^T \phi(x_j)$

$$\text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i^T x) + b\right)$$

对偶问题仍然是凸优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) - \sum_{k=1}^l \alpha_k$$

$$0 \leq \alpha_i \leq C$$

$$\sum_{j=1}^l \alpha_j y_j = 0$$

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i^T x_j) - \sum_{k=1}^l \alpha_k$$

$$0 \leq \alpha_i \leq C$$

$$\sum_{j=1}^l \alpha_j y_j = 0$$

Mercer条件

Mercer条件

对任意的有限个样本的样本集，核矩阵半正定

$$K_{ij} = k(x_i, x_j)$$

核函数	计算公式
线性核	$K(x_i, x_j) = x_i^T x_j$
多项式核	$K(x_i, x_j) = (\gamma x_i^T x_j + b)^d$
径向基函数核/高斯核	$K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$
sigmoid 核	$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + b)$

支持向量机能够转化为非线性问题，归功于核函数

核函数

14.14 本集总结

1. SVM基本情况
2. 线性分类
3. 分类间隔
4. SVM，引入线性可分的问题，凸优化问题，见了松弛变量，惩罚因子，slater条件。
5. 用和映射和核函数转化成非线性模型。
6. 在原问题上如何用KKT条件来解决。

在openCV上的应用

https://docs.opencv.org/master/d1/d69/tutorial_table_of_content_ml.html