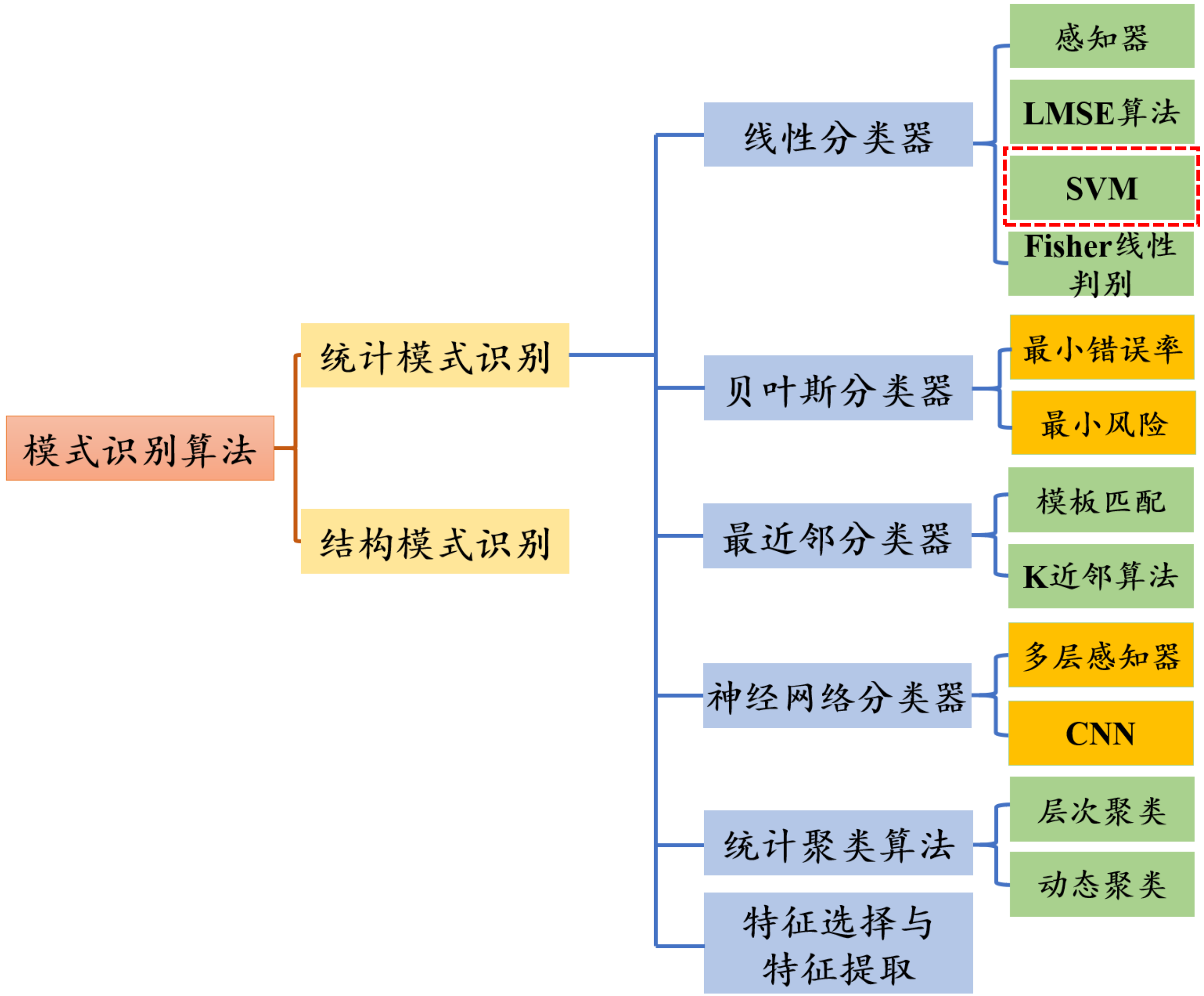
模式识别课程

支持向量机(SVM)

张俊超

中南大学
航空航天学院

模式识别-线性分类器


模式识别算法
统计模式识别
结构模式识别
线性分类器
贝叶斯分类器
最近邻分类器
神经网络分类器
统计聚类算法
特征选择与特征提取
感知器
LMSE算法
SVM
Fisher线性判别
最小错误率
最小风险
模板匹配
K近邻算法
多层感知器
CNN
层次聚类
动态聚类

- SVM学习要求：
  - 了解SVM原理
  - 借助SVM库，进行实际问题的求解

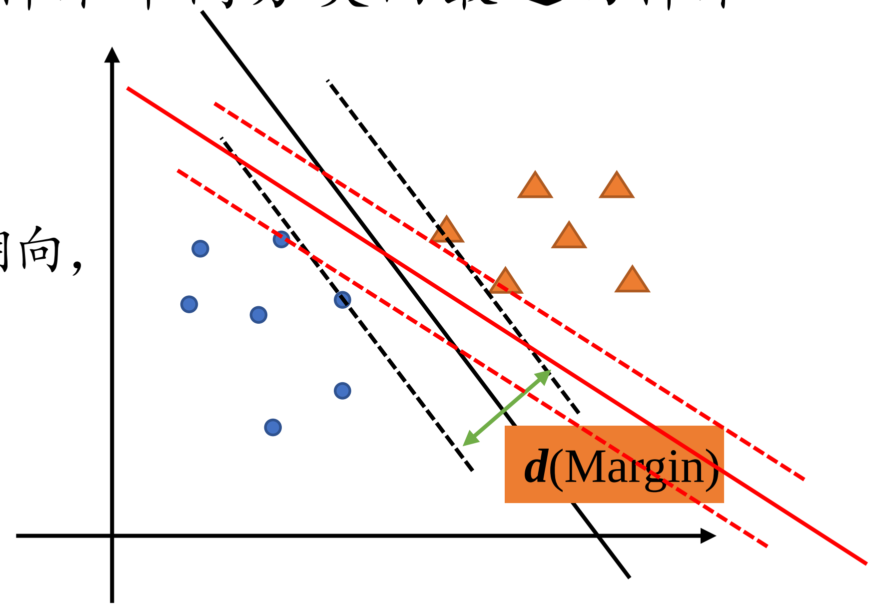
线性分类器 – SVM

概念：

- **最优分类超平面**：一个超平面，如果它能够将训练样本没有错误地分开，并且两类训练样本中离超平面最近的样本与超平面之间的距离是最大的。
- **分类间隔(Margin)**：两类样本中离分类面最近的样本到分类面的距离。
- **支持向量**

【不同的权向量，对应不同的边界朝向，分类间隔也不同。】

分类间隔是大还是小好呢？



线性分类器 -SVM

SVM的优化求解目标是：求取能带来最大分类间隔的权向量。

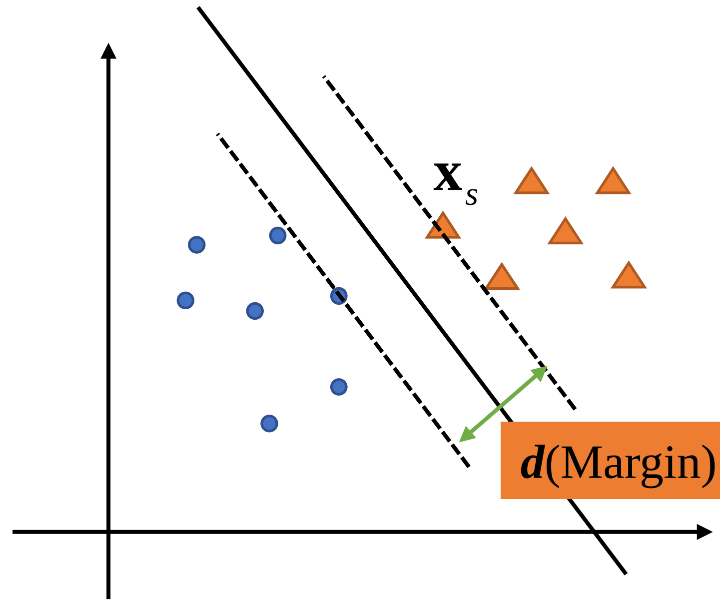
$$G(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$

样本到决策边界的距离为：

$$r = \frac{G(\mathbf{x})}{\|\mathbf{w}\|}$$

则(决策边界居中放置)：

$$d = 2\frac{G(\mathbf{x}_s)}{\|\mathbf{w}\|}$$



目标： 最大化d

线性分类器 –SVM

目标：最大化d

$$d = 2\frac{G(\mathbf{x}_s)}{\|\mathbf{w}\|}$$

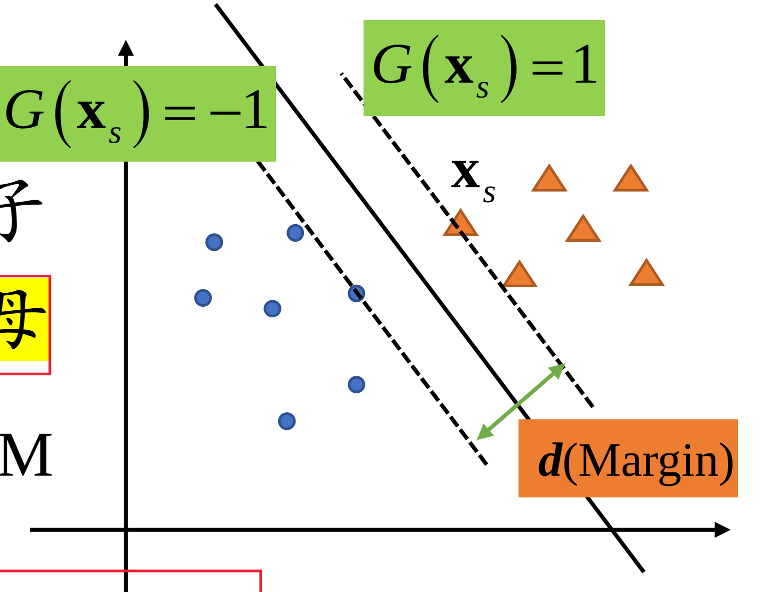
策略1：固定分母，最大化分子

策略2：固定分子，最小化分母

SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2, s.t. \left|G(\mathbf{x}_s)\right| = 1$$

二范数：平方和 开根号；



线性分类器－SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2$$
$$s.t. y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1, i = 1, 2, \ldots, n$$

其中 $y_i = \mathrm{sgn}(\mathbf{w}^T\mathbf{x}_i + w_0), y_i \in \{-1, 1\}$

这是一个凸优化的问题，可以用QP(quadratic program)的优化包直接求解。

Lagrange对偶变换，进行求解

线性分类器 –SVM

Lagrange对偶性： 引入拉格朗日对偶变量，便可以通过拉格朗日函数将约束条件融和到目标函数里去。

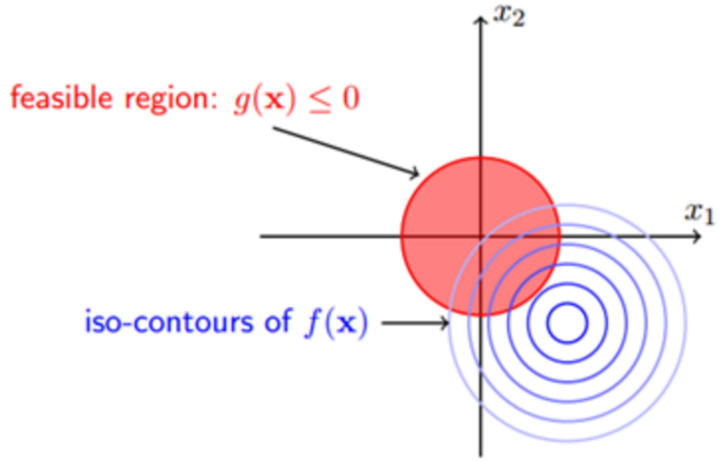
$$\min_x \; f(x)$$

$$\boxed{s.t. \;\; g(x) \le 0}$$

对应的 Lagrangian 与图形分别如下所示：

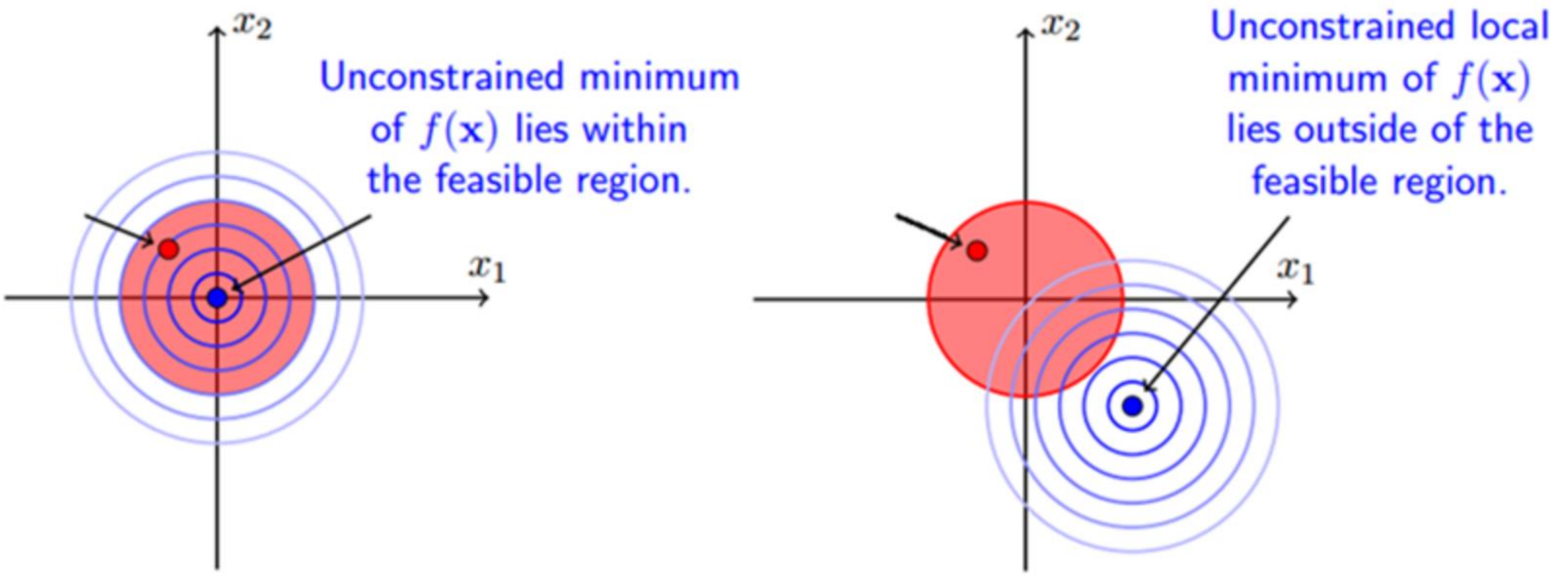
$$L(x, \lambda) = f(x) + \lambda g(x) \quad s.t. \lambda \ge 0$$

这时的可行解必须落在约束区域 $g(x)$ 之内，下图给出了目标函数的等高线与约束：



https://www.cnblogs.com/ooon/p/5721119.html

线性分类器 -SVM



由图可见可行解 x 只能在 $g(x) < 0$ 或者 $g(x) = 0$ 的区域里取得：

- 当可行解 x 落在 $g(x) < 0$ 的区域内，此时直接极小化 $f(x)$ 即可；即 $\lambda=0$
- 当可行解 x 落在 $g(x) = 0$ 即边界上，此时等价于等式约束优化问题.

综述所述，最优解满足 $\lambda g(x) = 0$

线性分类器-SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2,\quad s.t.\, y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \ge 1, i = 1,2,\ldots,n$$

$$\Downarrow$$

$$\boxed{\ell(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{n}\lambda_i\left(1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0)\right)}$$

记$g(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$，$\boxed{\text{则} g(\mathbf{w}) = \max\limits_{\lambda_i \ge 0} \ell(\mathbf{w}, w_0, \boldsymbol{\lambda})}$

【分析】$(1\text{-}y_i(\mathbf{w}^T\mathbf{x}_i + w_0)) \le 0$，$\lambda_i \ge 0 \rightarrow \boxed{\lambda_i(1\text{-}y_i(\mathbf{w}^T\mathbf{x}_i + w_0)) \le 0}$

$$\min_{\mathbf{w}, w_0} g(\mathbf{w}) = \min_{\mathbf{w}, w_0} \boxed{\max_{\lambda_i \ge 0} \ell(\mathbf{w}, w_0, \boldsymbol{\lambda})} = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{n}\lambda_i\left(1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0)\right)$$

线性分类器 – SVM

$$\min_{\mathbf{w},w_0}\max_{\lambda_i\ge 0}\ell(\mathbf{w},w_0,\boldsymbol{\lambda})=\frac{1}{2}\|\mathbf{w}\|^2+\sum_{i=1}^{n}\lambda_i\left(1-y_i(\mathbf{w}^T\mathbf{x}_i+w_0)\right)$$

$\Updownarrow$ 对偶问题　　**二者等价**

$$\max_{\lambda_i\ge 0}\min_{\mathbf{w},w_0}\ell(\mathbf{w},w_0,\boldsymbol{\lambda})=\frac{1}{2}\|\mathbf{w}\|^2+\sum_{i=1}^{n}\lambda_i\left(1-y_i(\mathbf{w}^T\mathbf{x}_i+w_0)\right)$$

强对偶： 目标函数与约束函数可微，任一对原始问题的解满足 KKT 条件。

1. $\frac{\partial\ell}{\partial\mathbf{w}}=0,\ \frac{\partial\ell}{\partial w_0}=0$
2. $\lambda_i\left(1-y_i(\mathbf{w}^T\mathbf{x}_i+w_0)\right)=0$
3. $\left(1-y_i(\mathbf{w}^T\mathbf{x}_i+w_0)\right)\le 0$
4. $\lambda_i\ge 0$

https://www.cnblogs.com/ooon/p/5723725.html

线性分类器 –SVM

$$\max_{\lambda_i \geq 0} \min_{\mathbf{w}, w_0} \ell(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \lambda_i \left(y_i (\mathbf{w}^T \mathbf{x}_i + w_0) - 1 \right)$$

Step 1:固定$\boldsymbol{\lambda}$,要让ℓ关于$\mathbf{w}$和w_0最小化:

$$\frac{\partial \ell}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i$$

$$\frac{\partial \ell}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\nabla_{\mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$$

$$\nabla_{\mathbf{x}} (\mathbf{y}^T \mathbf{x}) = \nabla_{\mathbf{x}} (\mathbf{x}^T \mathbf{y}) = \mathbf{y}$$

线性分类器－SVM

$$\ell(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \lambda_i \left(y_i (\mathbf{w}^T \mathbf{x}_i + w_0) - 1 \right)$$

$$= \frac{1}{2}\mathbf{w}^T\mathbf{w} - \mathbf{w}^T \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i - w_0 \sum_{i=1}^{n} \lambda_i y_i + \sum_{i=1}^{n} \lambda_i$$

$$= \frac{1}{2}\mathbf{w}^T \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i - \mathbf{w}^T \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i - w_0 \cdot 0 + \sum_{i=1}^{n} \lambda_i$$

$$= \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left(\sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i \right)^T \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i$$

$$\boxed{= \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j}$$

线性分类器－SVM

$Step2:$ 极大化$\boldsymbol{\lambda}$

$$\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}) = \left\{ \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}$$

$$s.t. \quad \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i y_i = 0, \quad i = 1, 2, \ldots, n$$

$Q(\cdot)$是对λ_i，$i = 1, 2, \ldots, n$的二次优化问题

SMO算法： *Sequential Minimal Optimization A Fast Algorithm for Training Support Vector Machines*(John C. Platt ,1998)

https://blog.csdn.net/techq/article/details/6171688 (Code)

本质： 一次更新两个变量，剩余的变量固定

线性分类器－SVM

$Step3:$ 求解$\mathbf{w}$和w_0

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}) = \left\{ \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}$$

$$s.t. \quad \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i y_i = 0, \quad i = 1, 2, \ldots, n$$

$$\Downarrow$$

$$\mathbf{w}^* = \sum_{i=1}^{n} \lambda_i^* y_i \mathbf{x}_i$$

$$w_0 = -\frac{1}{2}\left(\max_{i:y_i=-1} \left(\mathbf{w}^*\right)^T \mathbf{x}_i + \min_{i:y_i=1} \left(\mathbf{w}^*\right)^T \mathbf{x}_i \right)$$

线性分类器－SVM

Q:对分类超平面有影响的只有支持向量吗?

$$\max_{\lambda_i \geq 0} \ell(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \lambda_i \boxed{\left(y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1\right)}$$

对于非支持向量，红框部分大于0，$\lambda_i \geq 0$，为了满足最大化，λ_i=0 。故非支持向量不会影响超平面。

线性分类器-软间隔的SVM

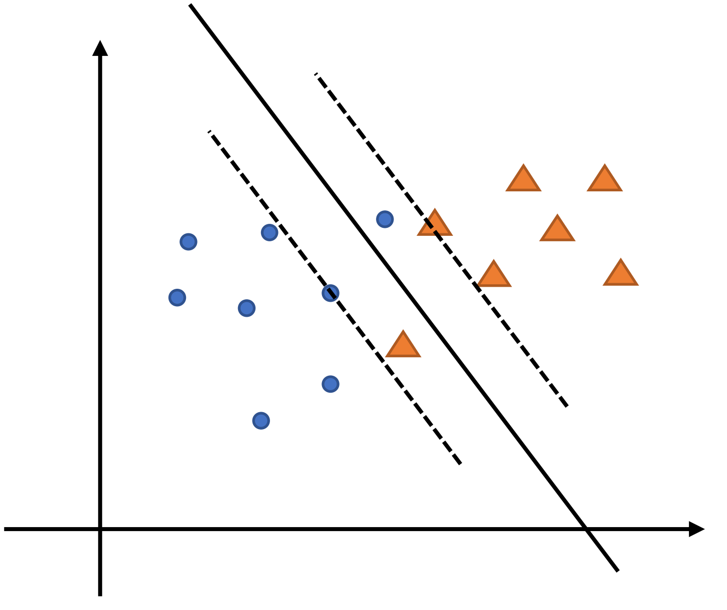
存在样本点满足$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) < 1$

引入松弛变量ξ_i,使$y_i(\mathbf{w}^T\mathbf{x}_i + w_0)-1+\xi_i \geq 0$成立

$$\min \frac{1}{2}\|\mathbf{w}\|^2$$
$$s.t. y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1, i = 1, 2, \ldots, n$$

$$\Downarrow$$

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t. y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1-\xi_i,$$
$$\xi_i \geq 0，\quad i = 1, 2, \ldots, n$$



线性分类器－软间隔的SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t. y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1-\xi_i,$$
$$\xi_i \geq 0，\ i = 1,2,\ldots,n$$

C:权衡“寻找间隔最大的超平面”(C小)和“保证数据点偏差量最小”(C大)

线性可分的情况，C的大小只影响中间过程，不影响最终结果，因为最终会$\sum_{i=1}^{n}\xi_i=0$

线性分类器－软间隔的SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t. y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1-\xi_i, \xi_i \geq 0, \quad i = 1,2,\ldots,n$$

$$\Downarrow$$

$$\min_{\mathbf{w},w_0} \max_{\lambda_i \geq 0, \ \gamma_i \geq 0} \ell(\mathbf{w}, w_0, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\lambda_i\left(y_i(\mathbf{w}^T\mathbf{x}_i + w_0) - 1 + \xi_i\right) - \sum_{i=1}^{N}\gamma_i\xi_i$$

Step 1:固定$\boldsymbol{\lambda}$和$\boldsymbol{\gamma}$,要让ℓ关于$\mathbf{w}$, w_0和ξ_i最小化:

$$\frac{\partial \ell}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}\lambda_i y_i \mathbf{x}_i$$

$$\frac{\partial \ell}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{n}\lambda_i y_i = 0$$

$$\frac{\partial \ell}{\partial \xi_i} = 0 \Rightarrow C - \lambda_i - \gamma_i = 0$$

线性分类器－软间隔的SVM

$Step2$:回代，极大化$\boldsymbol{\lambda}$

$Step3$:求解$\mathbf{w}$和w_0

原来：

$$\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}) = \left\{ \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}$$

$$s.t. \quad \lambda_i \ge 0, \sum_{i=1}^{n} \lambda_i y_i = 0, \quad i = 1, 2, \ldots, n$$

软间隔：

$$\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}) = \left\{ \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}$$

$$s.t. \quad \boxed{C \ge \lambda_i \ge 0,} \sum_{i=1}^{n} \lambda_i y_i = 0, \quad i = 1, 2, \ldots, n$$

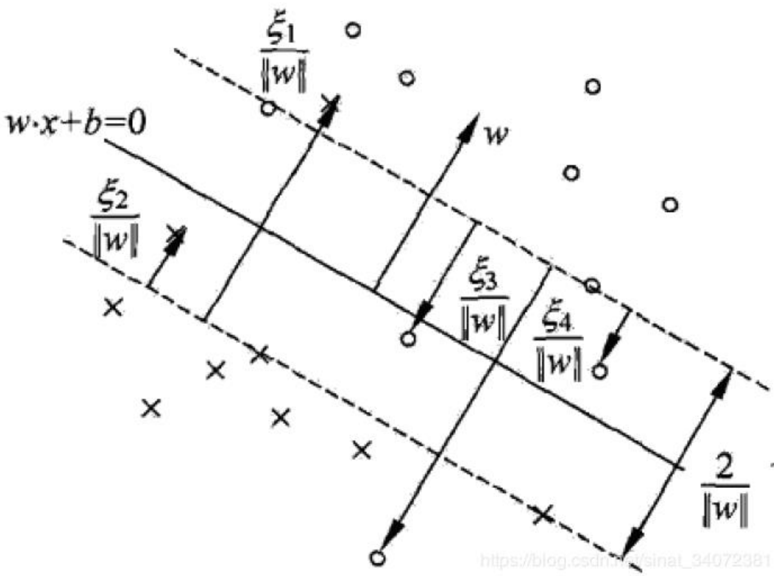
线性分类器－软间隔的SVM

【说明】

根据KKT条件得：

$$\lambda_i\left(y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1+\xi_i\right)=0$$

$$\gamma_i\xi_i=\left(C-\lambda_i\right)\xi_i=0$$



1.若λ_i=0，那么ξ_i=0，$y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1\geq 0$，即样本在间隔边界上或者已经被正确分类。

2.若0<$\lambda_i<C$，即$\gamma_i=C-\lambda_i>0$,那么ξ_i=0，$y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1$=0，即点在间隔边界上(支持向量)。

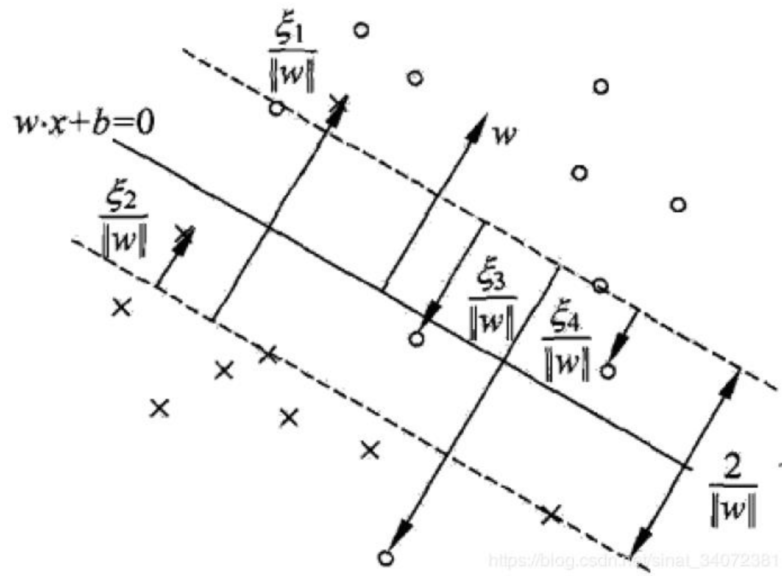
线性分类器－软间隔的SVM

【说明】

根据KKT条件得：

$$\lambda_i\left(y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1+\xi_i\right)=0$$

$$\gamma_i\xi_i=(C-\lambda_i)\xi_i=0$$



3.若λ_i=C，即$\gamma_i=C-\lambda_i=0$,那么$\xi_i\geq 0$

(1)若$0\leq\xi_i<1$,那么点被正确分类，但是却在超平面和自己的类别的间隔边界之间，比如图中的样本2和4。

(2)若$\xi_i=1$,那么点在超平面上，无法被正确分类(拒绝判断)。

(3)若$\xi_i>1$,那么点在超平面的另一侧,样本分类错误。

线性分类器－软间隔的SVM

【说明】

根据KKT条件得：

$$\lambda_i\left(y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1+\xi_i\right)=0$$

$$\gamma_i\xi_i=\left(C-\lambda_i\right)\xi_i=0$$

1.只有$\lambda_i=C$的样本才有$\xi_i>0$，其余的样本$\xi_i=0$

2.只有$y_i(\mathbf{w}^T\mathbf{x}_i+w_0)-1+\xi_i=0$的样本才有$\lambda_i>0$,其它$\lambda_i=0$

因此，可以选择$C>\lambda_i>0$的样本计算$\mathbf{w}$和w_0

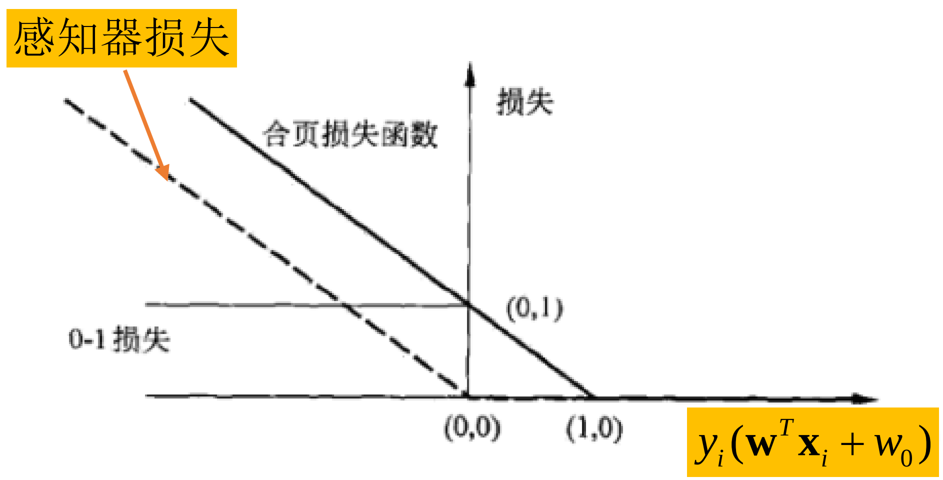
线性分类器－软间隔的SVM

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t. \; y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \ge 1 - \xi_i,$$
$$\xi_i \ge 0，\; i = 1, 2, \ldots, n$$

Hinge loss
(合页损失)

松弛变量：

$$\xi_i = \max\left(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0)\right)$$



线性分类器 –SVM

Matlab SVM Toolbox:

https://www.mathworks.com/help/stats/templatesvm.html

- **训练:**
  - templateSVM() (Create a Support Vector Machine template)
  - fitcsvm() (二分类SVM)
  - fitcecoc() (Fit a multiclass model for Support Vector Machine or other classifiers)
- **测试:**
  - predict()

```
t = templateSVM('KernelFunction','linear');
svm_model = fitcecoc(data_train,train_labels,'Learners',t);
Result = predict(svm_model,data_test);
```

线性分类器－SVM

线性分类器 –SVM

线性分类器 – 补充材料

SVM详解：

https://blog.csdn.net/macyang/article/details/38782399