2.9 带松弛因子的SVM：C-SVM

CSDN学院

2017年11月

大纲

- SVM基本原理
- 带松弛因子的SVM：C-SVM
- 核方法
- 支持向量回归
- Scikit learn中的SVM实现
- 案例分析

C-SVM

- 之前我们讨论的是样本可以完全线性可分的情况，即

$$y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right)\geq 1$$

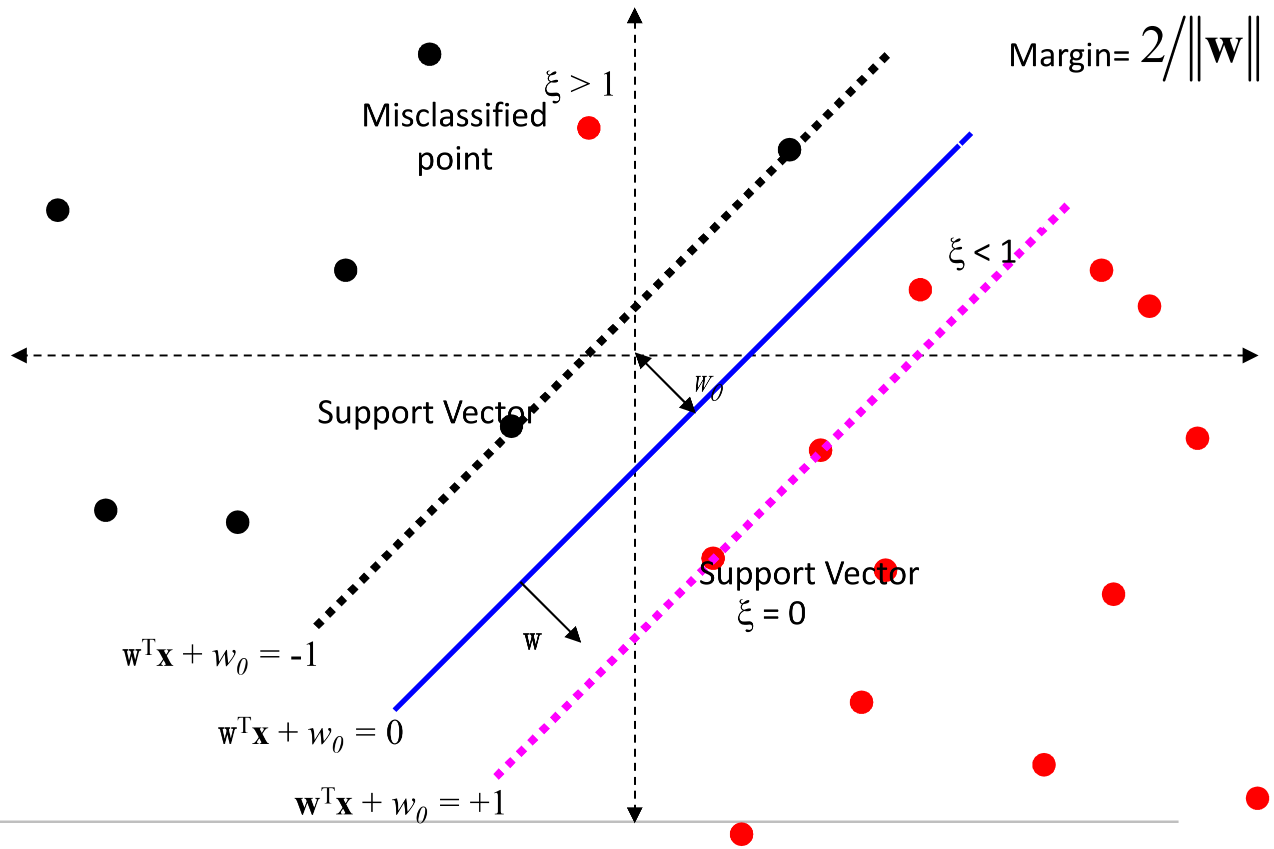
- 在实际问题中，数据不一定线性可分。因此解决方案引入软间隔（ soft margin ），允许一些样本出错，即允许某些样本不满足约束，将约束放松为

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

- 其中 ξ_i 称为松弛变量（ *slack variables* ）
- 但不满足约束的样本越少越好，因此增加松弛量限制

$$\xi_i\geq 0,\ \sum_i \xi_i\leq C$$

数据不完全线性可分


Margin= $2/\|\mathbf{w}\|$
$\xi > 1$
Misclassified point
$\xi < 1$
w_0
Support Vector
Support Vector
$\xi = 0$
w
$\mathbf{w}^T\mathbf{x} + w_0 = -1$
$\mathbf{w}^T\mathbf{x} + w_0 = 0$
$\mathbf{w}^T\mathbf{x} + w_0 = +1$

C-SVM

- 因此软间隔的SVM（C-SVM）目标函数为

$$\min \frac{1}{2}\|\mathbf{w}\|^2, \quad \text{subject to} \quad y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) \geq 1-\xi_i, \quad \forall i$$

- 且松弛变量 ξ_i 限制为 $\xi_i \geq 0, \sum_i \xi_i \leq C$ 。
- 可以将上公重写为

$$J(\mathbf{w})=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}\xi_i$$

$$\text{subject to} \quad y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) \geq 1-\xi_i,$$

$$\xi_i \geq 0$$

▶ C-SVM

CSDN 不止于代码

- 目标函数： $J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$

  subject to $\quad y_i\left(w_0 + \mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) \geq 1 - \xi_i,$

  $\xi_i \geq 0$

- 其中参数C控制间隔和松弛变量惩罚项之间的平衡
- 被误分的点的 $\xi_i > 1$ ，因此$\sum_i \xi_i$ 为被误分点的数目的上界，可视为训练误差
- 因此参数C可视为控制最小训练误差和模型复杂度的参数

▶ C-SVM——对偶

- 目标函数： $\min\left[\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right]$

  subject to $\quad y_i\left(w_0 + \mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) \geq 1 - \xi_i,$

  $\xi_i \geq 0$

- 对应的拉格朗日函数为

$$L(\boldsymbol{\alpha}, \boldsymbol{\mu}, w_0, \mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left(y_i\left(w_0 + \mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) - 1 + \xi_i\right) - \sum_{i=1}^{N}\mu_i\xi_i$$

- 其中 $\alpha_i \geq 0, \mu_i \geq 0$为拉格朗日乘子。

▶ C-SVM——对偶

- 对拉格朗日函数求偏导数，得

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w}^* = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \mu_i$$

► C-SVM——对偶

- 将结果代入拉格朗日函数，得到其对偶问题

$$Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle$$

- s.t. $\sum_{i=1}^{N}\alpha_i y_i=0$

  $0\le\alpha_i\le C$

- 与线性可分情况的不同之处仅在于约束为 $0\le\alpha_i\le C$
  - 线性可分情况 $:0\le\alpha_i$
- 最后的决策函数形式同线性可分情况

C-SVM – KKT Conditions

$$\alpha_i \geq 0$$

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

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

$$\mu_i \geq 0$$

$$\xi_i \geq 0$$

$$\mu_i \xi_i = 0$$

α的稀疏性

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

- 与线性可分情况类似，一些数据点 $\alpha_i=0$
  - 被正确分类，在支持超平面之外，对预测没有贡献
- 对 $\alpha_i>0$ 的点，必须满足

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

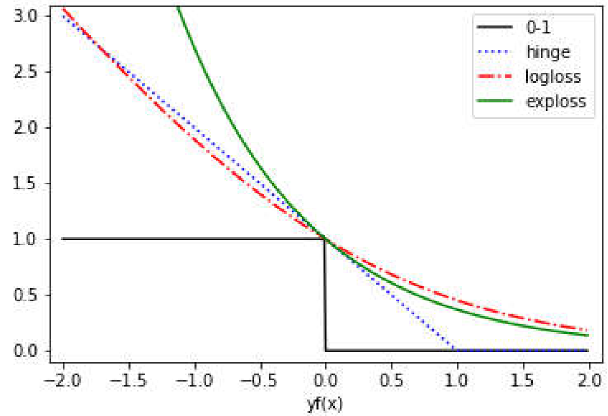
  - 若 $\alpha_i<C$，则 $\mu_i>0, \xi_i=0$，位于支持平面/边界上
  - 若 $\alpha_i=C$，则 $\xi_i\le 1$（位于支持平面/边界里面，或 $\xi_i>1$（被误分））

▶ 合页损失（hinge loss）

不止于代码

- 在C-SVM中，
  - 当 $y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right)\geq 1, \xi_i=0$
  - 其他点：$\xi_i=1-y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right)$
- 因此目标函数 $C\sum_{i=1}^{N}\xi_i+\frac{1}{2}\|\mathbf{w}\|^2$
- 可写成 $\sum_{i=1}^{N}\left[1-y_i\left(w_0+\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right)\right]_+ +\lambda\|\mathbf{w}\|^2=\sum_{i=1}^{N}\left[1-y_i\eta_i\right]_+ +\lambda\|\mathbf{w}\|^2$
- 其中λ起到*C*的作用。该损失函数称为Hinge Loss

▶ 分类任务中的损失函数



- 0-1损失函数：$l_{0/1}(y, f(x)) = \begin{cases} 1 & yf(x) < 0 \\ 0 & othereise \end{cases}$
  - 不连续，不易于优化计算
- Logistic损失（logloss）：$l_{\log}(y, f(x)) = \log(1 + \exp(-yf(x)))$
  - Logistic回归 / 多层感知机，可得到类别的概率
- 合页损失：$l_{hinge}(y, f(x)) = \max(0, 1 - yf(x))$
  - SVM、稀疏解、不能直接得到类别的概率

指数损失：$l_{\exp}(y, f(x)) = \exp(-yf(x))$

AdaBoost

► SVM：合页损失 + L2正则

- Recall：机器学习一般模型的目标函数为：

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{N} l\left(f\left(\mathbf{x}_i;\boldsymbol{\theta}\right), y_i\right) + \Omega\left(\boldsymbol{\theta}\right)$$

- 当损失函数取负log似然，正则项取L2时，得到L2正则的Logistic回归

$$J(\mathbf{w}) = \sum_{i=1}^{N} -\left[y_i \log\left(\mu_i\right) + \left(1-y_i\right)\log\left(1-\mu_i\right)\right] + \lambda\left\|\mathbf{w}\right\|_2^2$$

- 当损失函数取合页损失，正则项取L2正则，得到SVM：

$$J(\mathbf{w},\lambda) = \sum_{i=1}^{N} L\left(y_i, f\left(\mathbf{x}_i;\mathbf{w}\right)\right) + \lambda\left\|\mathbf{w}\right\|^2$$

THANK YOU