10/17/2019 syn

参考文档

了解SVM

Logistic回归

给定一些数据点,它们分别属于两个不同的类,现在要找到一个线性分类器把这些数据分成两类。如果用x表示数据点,用y表示类别(y可以取1或者-1,分别代表两个不同的类,一个线性分类器的学习目标便是要在n维的数据空间中找到一个超平面(hyper plane),这个超平面的方程可以表示为

[1]
$$w^T x + b = 0$$

Cost Function

https://yoyoyohamapi.gitbooks.io/mit-ml/content/SVM/articles/代价函数.html

[2]

$$\min_{ heta} rac{1}{m} [\sum_{i=1}^n y^i (-\log h_{ heta}(x^i)) + (1-y^i) (1-\log(1-h_{ heta}(x^i)))] + rac{1}{2} \sum_{j=1}^n heta_j^2$$

$$h_{ heta}(x) = rac{1}{1 + \exp^{- heta^T x}}$$

Large Margin Classifier

点到直线的距离

$$(x_0,y_0)$$
 到 $Ax+By+C=0$ 距离可用如下公式表示: $d=rac{|Ax_0+By_0+C|}{\sqrt{A^2+B^2}}$

then:

$$(x_0,y_0)$$
 到 $w^TX+b=0$ 距离可用如下公式表示: $d=rac{|wx_0+b|}{\sqrt{A^2+B^2}}$

Margin

1. 为了预测精度足够高, y=1时, 希望 θ Tx≥1, y=0时, 希望 θ Tx≤-1

10/17/2019

2. margin线上的点 (x_0,y_0) 到直线 $w^TX+b=0$ 的距离为 [3] $d=rac{1}{||w||}$,因为margin线为 $w^Tx+b=1$ 或 $w^Tx+b=-1$

3. 将margin线加入y, 即 $y(w^Tx+b)=-1$

问题求解变为

[4]

$$\max \frac{1}{||w||}$$

$$s.t.y_i(w^Tx_i+b) \geq 1$$

$$i = 1, ..., n$$

深入SVM

从线性可分到线性不可分

由于求的最大值相当于求的最小值,所以上述目标函数等价于[5]

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

$$s.t.y_i(w^Tx_i+b) \geq 1$$

$$i = 1, ..., n$$

最优解问题分类

- 1. 无约束问题优化,可通过求导
- 2. 有等式约束优化,可通过拉格朗日乘子法
- 3. 有不等式约束优化,可通过拉格朗日对偶性解决,满足KKT条件

拉格朗日对偶性

10/17/2019 s

我们的问题符合第三类最优解问题:

$$\ell(w,b,lpha) = rac{1}{2} ||w||^2 - \sum_{i=1}^n lpha_i (y_i(w^T x_{(i)} + b) - 1)$$

令:

$$\psi(w) = \max_{lpha_i \geq 0} \ell(w,b,lpha)$$

 \Rightarrow

$$\psi(w) = \min_{w,b} \max_{lpha_i \geq 0} \ell(w,b,lpha)$$

有兰格朗日对偶性:

$$P^* = \min_{w,b} \max_{lpha_i \geq 0} \ell(w,b,lpha)$$

$$D^* = \max_{lpha_i \geq 0} \min_{w,b} \ell(w,b,lpha)$$

$$D^* < P^*$$

在满足KKT条件,这两者相等,这个时候就可以通过求解对偶问题来间接地求解原始问题,KKT条件:

s.t.

$$g(x) \leq 0$$

$$h(x) \equiv 0$$

 \Rightarrow

$$\ell(x, \alpha, \lambda) = f(x) + \alpha g(x) + \lambda h(x)$$

s.t.

10/17/2019 svn

$$a_i \geq 0, g(x) \leq 0, \sum_{i=1}^n lpha_i g(x_i) = 0, \lambda
eq 0$$

此问题可变为:

[6]

$$\ell(w,b,lpha) = rac{1}{2} ||w||^2 + \sum_{i=1}^n lpha_i (1 - y_i(w^T x_{(i)} + b))$$

$$a_i \geq 0, g(x) = 1 - y_i(w^Tx_{(i)} + b) \leq 0, \sum_{i=1}^n lpha_i g(x_i) = 0$$

 $\ell(w,b,\alpha)$ 有最大值,因为 $a_i \geq 0, g(x) \leq 0$, 分别对w,b求导:

$$\frac{\partial \ell}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

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

带入方程,⇒

$$\ell(w,b,lpha) = rac{1}{2} ||w||^2 + \sum_{i=1}^n lpha_i (1 - y_i(w^T x_{(i)} + b))$$

$$\ell(w,b,lpha) = rac{1}{2} w^T w + \sum_{i=1}^n lpha_i - \sum_{i=1}^n lpha_i y_i w^t x_i - b \sum_{i=1}^n lpha_i y_i$$

$$\ell(w,b,lpha) = rac{1}{2} w^T w + \sum_{i=1}^n lpha_i - \sum_{i=1}^n lpha_i y_i w^T x_i$$

$$\ell(w,b,lpha) = rac{1}{2} (\sum_{i=1}^n lpha_i y_i x_i)^T (\sum_{i=1}^n lpha_i y_i x_i) + \sum_{i=1}^n lpha_i - \sum_{i=1}^n lpha_i y_i (\sum_{j=1}^n lpha_j y_j x_j)^T x_i$$

$$\ell(w,b,lpha) = rac{1}{2} \sum_{i=1,j=1}^n lpha_i lpha_j y_i y_j x_i^T x_j + \sum_{i=1}^n lpha_i - \sum_{i=1,j=1}^n lpha_i lpha_j y_i y_j x_i^T x_j$$

$$\ell(w,b,lpha) = \sum_{i=1}^n lpha_i - rac{1}{2} \sum_{i=1,j=1}^n lpha_i lpha_j y_i y_j x_i^T x_j$$

当 $\ell(w,b,lpha)$ 取到极值时[7]:

$$\ell(w,b,lpha) = \sum_{i=1}^n lpha_i - rac{1}{2} \sum_{i=1,j=1}^n lpha_i lpha_j y_i y_j x_i^T x_j = f(x)$$

核函数(mercer定理,半正定)

分类:

- 1. 线性核函数 $K(x_i,x_j)=x_i^Tx_j$
- 2. 多项式核函数 $K(x_i,x_j)=(\gamma_i^Tx_j+b)$
- 3. 高斯核函数 $K(x_i,x_j)=\exp(-\gamma||x_i-x_j||^2)$

10/17/2019 svn

4. 拉格拉斯核函数

5. sigmod核函数

假设:

$$x = egin{bmatrix} x_1 \ x_2 \end{bmatrix}$$

映射到

$$\phi = egin{bmatrix} x_1x_1 \ x_1x_2 \ x_2x_1 \ x_2x_2 \end{bmatrix}$$

 \Rightarrow

$$K(\phi(m),\phi(n)) = \phi(m)^T \phi(n)$$

$$K(m,n) = m_1 m_1 n_1 n_1 + m_1 m_2 n_1 n_2 + m_2 m_1 n_2 n_1 + m_2 m_2 n_2 n_2$$

$$K(m,n) = m_1 m_1 n_1 n_1 + 2 m_1 m_2 n_1 n_2 + m_2 m_2 n_2 n_2$$

$$K(m,n) = (m_1n_1 + m_2n_2)^2$$

$$K(m, n) = (m^T n)^2$$

SMO

取 α_1, α_2 作为参数

$$D^* = \max_{lpha_i \geq 0} \min_{w,b} \ell(w,b,lpha)$$

$$\ell(w,b,lpha) = \sum_{i=1}^n lpha_i - rac{1}{2} \sum_{i=1,j=1}^n lpha_i lpha_j y_i y_j x_i^T x_j = f(x)$$

$$\ell(w,b,lpha) = lpha_1 + lpha_2 + \sum_3^n lpha_i - rac{1}{2} [lpha_1 y_1 \sum_{j=1}^n lpha_j y_j (x_1^T x_j) + lpha_2 y_2 \sum_{j=1}^n lpha_j y_j (x_2^T x_j) + \sum_{i=3,j=1}^n lpha_i lpha_j y_i y_j (x_i^T x_j)] + \sum_{i=3}^n lpha_i$$

$$\begin{array}{l} \ell(w,b,\alpha) = \alpha_1 + \alpha_2 + \sum_3^n \alpha_i - \frac{1}{2} [\alpha_1 y_1 \alpha_1 y_1(x_1^T x_1) + \alpha_1 y_1 \alpha_2 y_2(x_1^T x_2) + \\ \alpha_1 y_1 \sum_{j=3}^n \alpha_j y_j(x_1^T x_j) + \alpha_2 y_2 \alpha_1 y_1(x_2^T x_1) + \alpha_2 y_2 \alpha_2 y_2(x_2^T x_2) + \\ \alpha_2 y_2 \sum_{j=3}^n \alpha_j y_j(x_2^T x_j) + \sum_{i=3,j=3}^n \alpha_i \alpha_j y_i y_j(x_i^T x_j)] \end{array}$$

$$\ell(w,b,lpha) = lpha_1 + lpha_2 - rac{1}{2} [lpha_1 y_1 lpha_1 y_1 (x_1^T x_1) + lpha_1 y_1 lpha_2 y_2 (x_1^T x_2) + lpha_1 y_1 \sum_{j=3}^n lpha_j y_j (x_1^T x_j) + lpha_2 y_2 lpha_1 y_1 (x_2^T x_1) + lpha_2 y_2 lpha_2 y_2 (x_2^T x_2) + lpha_2 y_2 \sum_{j=3}^n lpha_j y_j (x_2^T x_j] + C$$

10/17/2019 svm

$$\begin{array}{l} \ell(w,b,\alpha) = \alpha_1 + \alpha_2 - \frac{1}{2}[\alpha_1^2y_1^2(x_1^Tx_1) + \alpha_1y_1\alpha_2y_2(x_1^Tx_2) + \alpha_1y_1\sum_{j=3}^n\alpha_jy_j(x_1^Tx_j) + \\ \alpha_2y_2\alpha_1y_1(x_2^Tx_1) + \alpha_2^2y_2^2(x_2^Tx_2) + \alpha_2y_2\sum_{j=3}^n\alpha_jy_j(x_2^Tx_j] + C \end{array}$$

由于
$$y \in \{-1,1\}$$
,所以 $y^2 = 1$,由于线性核函数 $x_1^T x_2 = x_2^T x_1$

 \Rightarrow

$$\ell(w,b,lpha) = lpha_1 + lpha_2 - rac{1}{2} [lpha_1^2(x_1^Tx_1) + 2lpha_1y_1lpha_2y_2(x_1^Tx_2) + lpha_1y_1\sum_{j=3}^nlpha_jy_j(x_1^Tx_j) + lpha_2^2(x_2^Tx_2) + lpha_2y_2\sum_{j=3}^nlpha_jy_j(x_2^Tx_j] + C$$

由
$$\Rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$
,所以

$$\alpha_1 y_1 + \alpha_2 y_2 + \sum_{i=3}^n \alpha_i y_i = 0$$

 \Rightarrow

$$egin{aligned} lpha_1 y_1 + lpha_2 y_2 &= \zeta \ lpha_1^{old} y_1 + lpha_2^{old} y_2 &= lpha_1^{new} y_1 + lpha_2^{new} y_2 &= \zeta \end{aligned}$$

 \Rightarrow

$$lpha_1^{old}y_1y_1+lpha_2^{old}y_2y_1=\zeta y_1$$

 \Rightarrow

$$lpha_1^{old} = \zeta y_1 - lpha_2^{old} y_2 y_1$$

 \Rightarrow

$$lpha_1^{old}=\zeta^{'}-slpha_2^{old}$$

 \Rightarrow

$$\ell(w,b,lpha) = lpha_1 + lpha_2 - rac{1}{2} [lpha_1^2(x_1^Tx_1) + 2lpha_1y_1lpha_2y_2(x_1^Tx_2) + lpha_1y_1\sum_{j=3}^nlpha_jy_j(x_1^Tx_j) + lpha_2^2(x_2^Tx_2) + lpha_2y_2\sum_{j=3}^nlpha_jy_j(x_2^Tx_j] + C$$

 \Rightarrow

$$\ell(w,b,lpha) = (\zeta^{'} - slpha_2) + lpha_2 - rac{1}{2}[(\zeta^{'} - slpha_2)^2(x_1^Tx_1) + 2(\zeta^{'} - slpha_2)y_1lpha_2y_2(x_1^Tx_2) + (\zeta^{'} - slpha_2)y_1\sum_{j=3}^{n}lpha_jy_j(x_1^Tx_j) + lpha_2^2(x_2^Tx_2) + lpha_2y_2\sum_{j=3}^{n}lpha_jy_j(x_2^Tx_j] + C$$

函数中变量只有 α_2 , 求取最大值, 需要对 α_2 求偏导数:

$$\frac{\partial \ell}{\partial \alpha_2} = 0$$

 \Rightarrow

$$lpha_{2}=C^{'}$$

10/17/2019 svm

 \Rightarrow

$$lpha_{1}=\zeta^{'}-slpha_{2}$$

 \Rightarrow

$$lpha_2^{new} = C^{''}$$

 \Rightarrow

$$lpha_1^{new} = C^{'''}$$