SVM 分类器

机器学习笔记 create by siwanghu v1.0

SVM 分类器，解决二分类问题

设训练集为$X=\{(x_1, y_1), (x_2, y_2), (x_i, y_i), \ldots, (x_n, y_n)\}$，$x_i$是 n 维向量，代表着样本有 n 个特征。$y_i \epsilon \{-1,1\}$，表示是一个二分类的问题，我们需要在空间中求得一条划分平面$wx+b=0$可以将训练集中的样本点划分为两类。

函数距离：$y_i(wx_i+b)$

几何距离：$\frac{|y_i(wx_i+b)|}{\sqrt{w^2}}$

由于几何距离会因为w和 b 同等系数改变而改变，实际上划分平面并没有变，所以我们采用函数距离来定义样本点到划分平面的距离。

为了约束距离，我们假设样本一侧最近的点到划分平面的函数距离为 1，这样做是为了方便优化，不会影响最终的计算。

SVM 约束条件：

隔离带$\frac{2}{\sqrt{w^2}}$的几何距离最大，等价于$\frac{1}{2}w^2$最小

训练集中的每个样本函数距离$y_i(wx_i+b)>1$

硬间隔 SVM 推导

SVM 约束：

min：$\frac{1}{2}w^2$

St：$y_i(wx_i+b)>1$

采用不等式的朗格朗日乘子法：（约束条件中，不等式必须要小于等于 0）

构造朗格朗日函数(有多少样本点，就有多少α_i)：

$$L=\frac{1}{2}w^2+\alpha_i\sum[1-y_i(wx_i+b)]$$

对 w 和 b 求偏导：

$$\frac{\partial L}{\partial w} = \mathrm{w} - \sum \alpha_{\mathrm{i}} \mathrm{y_i} \mathrm{x_i}$$

$$\frac{\partial L}{\partial \mathrm{b}} = -\sum \alpha_{\mathrm{i}} \mathrm{y_i}$$

令偏导等于 0，所以：

$$\mathrm{w} = \sum \alpha_{\mathrm{i}} \mathrm{y_i} \mathrm{x_i}$$

$$\sum \alpha_{\mathrm{i}} \mathrm{y_i} = 0$$

将这两个式子带入构造的朗格朗日函数 L 中，将消去 w 和 b，只会保留α_{i}：

$$\mathrm{L} = \frac{1}{2} \sum \sum \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} y_{\mathrm{i}} y_j x_{\mathrm{i}} x_j + \sum \alpha_{\mathrm{i}} - \sum \sum \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} y_{\mathrm{i}} y_j x_{\mathrm{i}} x_j - \sum \alpha_{\mathrm{i}} y_{\mathrm{i}} \mathrm{b}$$

有应为：$\sum \alpha_{\mathrm{i}} \mathrm{y_i} = 0$，所以：

$$\mathrm{L} = \sum \alpha_{\mathrm{i}} - \frac{1}{2} \sum \sum \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} y_{\mathrm{i}} y_j x_{\mathrm{i}} x_j$$

变成了对偶问题，求满足上式的α_{i}，使得L最大，也可以求$-$L的最小，最终约束为：

$$\min: \frac{1}{2} \sum \sum \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} y_{\mathrm{i}} y_j x_{\mathrm{i}} x_j - \sum \alpha_{\mathrm{i}}$$

$$\mathrm{st:} \quad \sum \alpha_{\mathrm{i}} \mathrm{y_i} = 0$$

$$\alpha_{\mathrm{i}} \geq 0$$

对上面这个问题，求出每个α_{i}，直接采用等式的拉格朗日解法求即可，如果存在$\alpha_{\mathrm{i}} \leq 0$，表示函数内部没有最优解，最优解在边缘处，令那些$\alpha_{\mathrm{i}} \leq 0$的变成$\alpha_{\mathrm{i}} = 0$，遍历猜测求出最优解和每个$\alpha_{\mathrm{i}}$。

带入$\mathrm{w} = \sum \alpha_{\mathrm{i}} \mathrm{y_i} \mathrm{x_i}$便可以求出 w，至于 b 的求解，非常的巧妙，在解出每个α_{i}后，那些$\alpha_{\mathrm{i}} \geq 0$的点对应的样本正好落在隔离带边缘（隔离带的中间便是我们要求的划分平面），我们称之为支持向量，利用支持向量，便可以求出 b。

最终求出划分平面$\mathrm{wx} + \mathrm{b} = 0$

软间隔 SVM 推导

软间隔 SVM 的意思是在隔离带的中间，我们允许少数的点落在隔离带中，软间隔 SVM 可以解决一部分的非线性划分问题，要求样本中的大部分点是二类线性可划分，少部分点无法线性划分，软间隔 SVM 其实还可以在一定程度上解决硬间隔 SVM 过拟合的问题。

这样增加一个变量$\varepsilon_{\mathrm{i}} \geq 0$，每个样本都要增加这个变量，意思是函数间隔可以不大于 1，让它大于$1-\varepsilon_{\mathrm{i}}$，起到减小硬间隔约束条件。

软间隔 SVM 约束(C 越大对误分类的惩罚越大，越小对误分类的惩罚越小，是自定义的参数)：

min: $\frac{1}{2}\mathrm{w}^2+\mathrm{C}\sum \varepsilon_{\mathrm{i}}$

st: $\mathrm{y_i}(\mathrm{w}\mathrm{x_i}+\mathrm{b}) \geq 1-\varepsilon_{\mathrm{i}}$

$\varepsilon_{\mathrm{i}} \geq 0$

采用不等式的朗格朗日乘子法：（约束条件中，不等式必须要小于等于 0）

构造朗格朗日函数(有多少样本点，就有多少α_{i}，ε_{i}，二者是朗格朗日乘子)：

$$\mathrm{L}=\frac{1}{2}\mathrm{w}^2+\mathrm{C}\sum \varepsilon_{\mathrm{i}}+\alpha_{\mathrm{i}}\sum[1-\varepsilon_{\mathrm{i}}-\mathrm{y_i}(\mathrm{w}\mathrm{x_i}+\mathrm{b})]-\mu_{\mathrm{i}}\sum \varepsilon_{\mathrm{i}}$$

对 w，b，ε_{i}求偏导：

$$\frac{\partial L}{\partial w}=\mathrm{w}-\sum \alpha_{\mathrm{i}}\,\mathrm{y_i}\mathrm{x_i}$$

$$\frac{\partial L}{\partial b}=-\sum \alpha_{\mathrm{i}}\,\mathrm{y_i}$$

$$\frac{\partial L}{\partial \varepsilon_{\mathrm{i}}}=C-\alpha_{\mathrm{i}}-\mu_{\mathrm{i}}$$

令它们的导数都为 0，所以：

$\mathrm{w}=\sum \alpha_{\mathrm{i}}\,\mathrm{y_i}\mathrm{x_i}$

$\sum \alpha_{\mathrm{i}}\,\mathrm{y_i}=0$

$C=\alpha_{\mathrm{i}}+\mu_{\mathrm{i}}$

将上述式子带入构造的朗格朗日函数 L 中，所以

$$\mathrm{L}=\frac{1}{2}\sum\sum \alpha_{\mathrm{i}}\alpha_{\mathrm{j}}y_{\mathrm{i}}y_j x_{\mathrm{i}}x_j+\mathrm{C}\sum \varepsilon_{\mathrm{i}}+\sum \alpha_{\mathrm{i}}-\sum \alpha_{\mathrm{i}}\,\varepsilon_{\mathrm{i}}-\sum\sum \alpha_{\mathrm{i}}\alpha_{\mathrm{j}}y_{\mathrm{i}}y_j x_{\mathrm{i}}x_j-\sum \alpha_{\mathrm{i}}y_{\mathrm{i}}\,\mathrm{b}-\mu_{\mathrm{i}}\sum \varepsilon_{\mathrm{i}}$$

因为$\sum\alpha_i y_i=0$，$C=\alpha_i+\mu_i$，对上述式子 L 化简：

$$L=-\frac{1}{2}\sum\sum\alpha_i\alpha_j y_i y_j x_i x_j+C\sum\varepsilon_i+\sum\alpha_i-\sum\alpha_i\varepsilon_i-\mu_i\sum\varepsilon_i$$

$$=-\frac{1}{2}\sum\sum\alpha_i\alpha_j y_i y_j x_i x_j+C\sum\varepsilon_i+\sum\alpha_i-\varepsilon_i\sum(\alpha_i+\mu_i)$$

$$=-\frac{1}{2}\sum\sum\alpha_i\alpha_j y_i y_j x_i x_j+\sum\alpha_i$$

化简之后发现软间隔 SVM 的对偶约简与硬间隔 SVM 的对偶约简相同

变成了对偶问题，求满足上式的α_i，使得L最大，也可以求$-$L的最小，最终约束为：

$$\min: \frac{1}{2}\sum\sum\alpha_i\alpha_j y_i y_j x_i x_j-\sum\alpha_i$$

$$\text{st:}\quad \sum\alpha_i y_i=0$$

$$0\le\alpha_i\le C$$

求出α_i，利用$w=\sum\alpha_i y_i x_i$，算出 w

随便选择一个$0\le\alpha_j\le C$，$b=y_j-\sum y_i\alpha_i x_i x_j$

最终求出划分平面$wx+b=0$

核技巧（并不复杂）

我们知道如果在低维空间中，用软间隔 SVM 也不能对样本二分类，我们可以将低维空间中的样本经过一个函数投影到高维空间中，如果将低维空间中的样本全部都投影到高维空间中的，样本在高维空间可线性划分。

然而，**我们不容易找到这个映射函数。**描述几何空间中的样本分别，我们常用任意两个样本点在空间中的距离和角度描述，这两个的计算（高中就学过）都需要计算两个向量的点积，**也就是说任意两个样本点之间的点积可以很好的代表样本在空间中的分布。**

核函数起到这样的作用，它使用低维空间的向量，将低维空间中两个向量带入核函数后，可以算出这两个向量投射到高维空间中后向量点积的运算结果。

观察上面 SVM 的对偶问题：

$$\min: \frac{1}{2}\sum\sum\alpha_i\alpha_j y_i y_j x_i x_j-\sum\alpha_i$$

$$\text{st:}\quad \sum\alpha_i y_i=0$$

$$0\le\alpha_i\le C$$

转成：

$$\min: \frac{1}{2}\sum\sum \alpha_{\mathrm{i}}\alpha_{\mathrm{j}} y_{\mathrm{i}} y_j \mathrm{K}(x_{\mathrm{i}}, x_j) - \sum \alpha_{\mathrm{i}}$$

$$\text{st:} \quad \sum \alpha_{\mathrm{i}} \mathrm{y}_{\mathrm{i}} = 0$$

$$0 \le \alpha_{\mathrm{i}} \le C$$

其中$\mathbf{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$代表采用核函数 K 将$\boldsymbol{x_i}, \boldsymbol{x_j}$计算出在高维空间中的向量点积。

什么函数可以作为核函数，我们有这样的约束，对于函数 K，将任意两个x_{i}, x_j带入函数 K 中，组成矩阵，$K_{i,j} = \begin{bmatrix} \mathrm{K}(x_1, x_1) & \dots & \mathrm{K}(x_1, x_n) \\ \dots & \dots & \dots \\ \mathrm{K}(x_n, x_1) & \dots & \mathrm{K}(x_n, x_n) \end{bmatrix}$，如果这个矩阵是半正定矩阵（矩阵所有特征值都大于等于 0），则函数 K 可以作为核函数。

常见的核函数有：线性核函数，多项式核函数，高斯核函数，拉普拉斯核函数，sigmoid 核函数。