法律声明

本课件包括演示文稿、示例、代码、题库、视频和声音等内容，深度之眼和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920

公众号

微 信

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

deepshare.net
深度之眼

支持向量机公式推导

Derivation of SVM

本节大纲

Outline

先修内容：西瓜书6.1、6.2、6.3

1.超平面的基本知识

2.SVM主问题

3.拉格朗日对偶

4.SVM对偶问题

超平面的基本知识

Basic knowledge of hyperplane

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

- 法向量恒垂直于超平面；
- 和法向量方向相同的点代入超平面方程恒大于等于0，否则恒小于等于0；
- 法向量和位移项唯一确定一个超平面；
- 等倍缩放法向量和位移项超平面不变；

点到超平面距离公式推导

deepshare.net
深度之眼

Derivation of distance formula from point to hyperplane

$$r=\frac{\left|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b\right|}{\|\boldsymbol{w}\|}$$

设点 $\boldsymbol{x}_0$ 到超平面的距离为 r，在超平面上的投影为 $\boldsymbol{x}_1$，那么法向量 $\boldsymbol{w}$ 一定与向量 $\overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_0}$ 平行,此时若考虑求两个向量内积的绝对值可得

$$\left|\boldsymbol{w}^{\mathrm{T}}\cdot\overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_0}\right|=\|\boldsymbol{w}\|\cdot|\pm 1|\cdot\left\|\overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_0}\right\|=\|\boldsymbol{w}\|\cdot r$$

又因为

$$\begin{aligned}
\boldsymbol{w}^{\mathrm{T}}\cdot\overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_0} &= w_1(x_0^1-x_1^1)+w_2(x_0^2-x_1^2)+\ldots+w_n(x_0^n-x_1^n)\\
&= w_1x_0^1+w_2x_0^2+\ldots+w_nx_0^n-(w_1x_1^1+w_2x_1^2+\ldots+w_nx_1^n)\\
&= w_1x_0^1+w_2x_0^2+\ldots+w_nx_0^n-(-b)\\
&= w_1x_0^1+w_2x_0^2+\ldots+w_nx_0^n+b=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_0+b
\end{aligned}$$

点到超平面距离公式推导

Derivation of distance formula from point to hyperplane

所以

$$\left|\boldsymbol{w}^{\mathrm{T}} \cdot \overrightarrow{\boldsymbol{x}_1 \boldsymbol{x}_0}\right| = \left|\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_0 + b\right| = \|\boldsymbol{w}\| \cdot r$$

$$r = \frac{\left|\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + b\right|}{\|\boldsymbol{w}\|}$$

SVM主问题

Primal problem of SVM

函数间隔：给定数据集 $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_m,y_m)\},y_i\in\{-1,+1\}$ 和超平面 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=0$，定义超平面关于样本 $(\boldsymbol{x}_i,y_i)$ 的函数间隔为

$$\hat{\gamma}_i=y_i(\boldsymbol{w}\boldsymbol{x}_i+b)$$

定义超平面关于数据集的函数间隔为超平面关于数据集中所有样本的函数间隔最小值，也即

$$\hat{\gamma}=\min_{1,2,\ldots,m}\hat{\gamma}_i$$

SVM主问题

Primal problem of SVM

deepshare.net
深度之眼

几何间隔：定义超平面关于样本 $(\boldsymbol{x}_i, y_i)$ 的几何间隔为

$$\gamma_i = \frac{\hat{\gamma}_i}{\|\boldsymbol{w}\|}$$

定义超平面关于数据集的几何间隔为超平面关于数据集中所有样本的几何间隔最小值，也即

$$\gamma = \min_{1,2,\ldots,m} \gamma_i = \frac{\hat{\gamma}}{\|\boldsymbol{w}\|}$$

SVM主问题

Primal problem of SVM

SVM的核心思想：求一个与已知数据集几何间隔最大的那个超平面，也即

$$
\begin{aligned}
\min \quad & \gamma \\
\text{s.t.} \quad & \gamma_i \geq \gamma, \quad i = 1, 2, \ldots, m
\end{aligned}
$$

$$
\begin{aligned}
\min \quad & \frac{\hat{\gamma}}{\|\boldsymbol{w}\|} \\
\text{s.t.} \quad & \frac{\hat{\gamma}_i}{\|\boldsymbol{w}\|} \geq \frac{\hat{\gamma}}{\|\boldsymbol{w}\|}, \quad i = 1, 2, \ldots, m
\end{aligned}
$$

$$
\begin{aligned}
\min \quad & \frac{\hat{\gamma}}{\|\boldsymbol{w}\|} \\
\text{s.t.} \quad & y_i(\boldsymbol{w}\boldsymbol{x}_i + b) \geq \hat{\gamma}, \quad i = 1, 2, \ldots, m
\end{aligned}
$$

$$
\begin{aligned}
\min \quad & \frac{1}{\|\boldsymbol{w}\|} \\
\text{s.t.} \quad & y_i(\boldsymbol{w}\boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \ldots, m
\end{aligned}
$$

SVM主问题

Primal problem of SVM

$$
\begin{aligned}
\min_{\boldsymbol{w}, b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\
\text{s.t.} \quad & 1 - y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \leqslant 0, \quad i = 1, 2, \ldots, m
\end{aligned}
$$

此优化问题本身就是一个很容易求解的凸优化问题，为啥非要用拉格朗日对偶去求解呢?

- 转化为对偶问题以后，对偶问题也很容易求解，同时还能很自然地引入核函数;
- 对偶问题的求解复杂度跟样本个数m成正比，主问题的求解复杂度跟参数维度呈正比，所以当样本个数远小于参数维度的时候用拉格朗日对偶求解效率更高;

dual : *bool, (default=True)*

Select the algorithm to either solve the dual or primal optimization problem. Prefer dual=False when n_samples > n_features.

https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC

拉格朗日对偶

Lagrange duality

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0 \quad (i = 1, \ldots, m) \\ & h_j(\boldsymbol{x}) = 0 \quad (j = 1, \ldots, n) \end{aligned}$$

设定义域为：$D = \boldsymbol{dom} f \cap \bigcap_{i=1}^{m} \boldsymbol{dom} g_i \cap \bigcap_{j=1}^{n} \boldsymbol{dom} h_j$ 可行集为：$\tilde{D} = \{\boldsymbol{x} | \boldsymbol{x} \in D, g_i(\boldsymbol{x}) \leq 0, h_j(\boldsymbol{x}) = 0\}$

拉格朗日对偶函数（以拉格朗日乘子为自变量的函数）：

$$\Gamma(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \in D} L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \in D} \left(f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{n} \lambda_j h_j(\boldsymbol{x}) \right)$$

- 拉格朗日对偶函数恒为凹函数【证明参见（王书宁 译.《凸优化》）3.2.3】；
- 当 $\boldsymbol{\mu} \succeq 0$ 时，拉格朗日对偶函数构成了优化问题最优值的下界；

拉格朗日对偶

Lagrange duality

deepshare.net 深度之眼

证明：因为当 $\boldsymbol{\mu} \succeq 0$ 时，对于可行集中的点 $\tilde{\boldsymbol{x}} \in \tilde{D}$ 下式恒成立

$$\sum_{i=1}^{n} \mu_i g_i(\tilde{\boldsymbol{x}}) + \sum_{j=1}^{m} \lambda_j h_j(\tilde{\boldsymbol{x}}) \leq 0$$

所以

$$\Gamma(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \in D} L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq L(\tilde{\boldsymbol{x}}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq f(\tilde{\boldsymbol{x}})$$

$$\Gamma(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq \min\{f(\tilde{\boldsymbol{x}})\} = p^*$$

拉格朗日对偶

Lagrange duality

deepshare.net 深度之眼

拉格朗日对偶问题：求对偶函数最大值的优化问题为拉格朗日对偶问题，也即

$$\begin{aligned} \max_{\boldsymbol{\mu},\boldsymbol{\lambda}} \quad & \Gamma(\boldsymbol{\mu},\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \succeq 0 \end{aligned}$$

设该优化问题的最优值为 d^*,显然 $d^* \leq p^*$,此时称为"弱对偶性"成立，若 $d^* = p^*$ 则称为"强对偶性"成立。

Slater条件：若主问题是凸优化问题，且可行域中存在一点能使得所有不等式约束的不等号成立，则强对偶性成立。

SVM对偶问题

Dual problem of SVM

deepshare.net
深度之眼

SVM的对偶函数：

$$\Gamma(\boldsymbol{\mu},\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\in D} L(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\in D}\left(f(\boldsymbol{x}) + \sum_{i=1}^{m}\mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{n}\lambda_j h_j(\boldsymbol{x})\right)$$

$$= \inf_{\boldsymbol{w},b\in D}\left(\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^{m}\alpha_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b))\right)$$

$$= \inf_{\boldsymbol{w},b\in D}\left(\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\alpha_i y_i\boldsymbol{w}^T\boldsymbol{x}_i - \sum_{i=1}^{m}\alpha_i y_i b\right)$$

$$= \inf_{\boldsymbol{w},b\in D}\left(\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\alpha_i y_i\boldsymbol{w}^T\boldsymbol{x}_i - b\sum_{i=1}^{m}\alpha_i y_i\right)$$

$$= \begin{cases} \inf_{\boldsymbol{w}\in D}\left\{\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\alpha_i y_i\boldsymbol{w}^T\boldsymbol{x}_i\right\}, & \text{if } \sum_{i=1}^{m}\alpha_i y_i = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

SVM对偶问题

Dual problem of SVM

SVM的对偶问题：

$$
\begin{aligned}
\max_{\boldsymbol{\mu},\boldsymbol{\lambda}} \quad & \Gamma(\boldsymbol{\mu},\boldsymbol{\lambda}) \\
\text{s.t.} \quad & \boldsymbol{\mu} \succeq 0
\end{aligned}
$$

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \begin{cases} \inf\limits_{\boldsymbol{w}\in D} \left\{\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i y_i \boldsymbol{w}^T \boldsymbol{x}_i\right\}, & \text{if } \sum_{i=1}^m \alpha_i y_i = 0 \\ -\infty, & \text{otherwise} \end{cases} \\
\text{s.t.} \quad & \boldsymbol{\alpha} \succeq 0
\end{aligned}
$$

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \inf_{\boldsymbol{w}\in D} \left\{\frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i y_i \boldsymbol{w}^T \boldsymbol{x}_i\right\} \\
\text{s.t.} \quad & \boldsymbol{\alpha} \succeq 0 \\
& \sum_{i=1}^m \alpha_i y_i = 0
\end{aligned}
$$

SVM对偶问题

Dual problem of SVM

其中

$$\inf_{\boldsymbol{w}\in D}\left\{\frac{1}{2}||\boldsymbol{w}||^2+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{w}^T\boldsymbol{x}_i\right\}$$

$$=\min_{\boldsymbol{w}\in D}\left\{\frac{1}{2}||\boldsymbol{w}||^2+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{w}^T\boldsymbol{x}_i\right\}$$

$$=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

所以，SVM的对偶问题具体形式为

$$\begin{aligned}\max_{\boldsymbol{\alpha}}\quad & \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\\ \text{s.t.}\quad & \boldsymbol{\alpha}\succeq 0\\ & \sum_{i=1}^{m}\alpha_i y_i=0\end{aligned}$$

此即为式6.11

SVM对偶问题

deepshare.net
深度之眼

Dual problem of SVM

KKT条件（必要条件）：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0 \quad (i = 1, \ldots, m) \\ & h_j(\boldsymbol{x}) = 0 \quad (j = 1, \ldots, n) \end{aligned}$$

设 $f(\boldsymbol{x}), g_i(\boldsymbol{x}), h_j(\boldsymbol{x})$ 具有连续的一阶偏导数，$\boldsymbol{x}^*$是局部问题的最优解且满足约束限制条件，那么一定存在 $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, \ldots, \mu_m^*)^T, \boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)^T$ 使得：

$$\begin{cases} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{n} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = 0 & (1) \\ h_j(\boldsymbol{x}^*) = 0 & (2) \\ g_i(\boldsymbol{x}^*) \le 0 & (3) \\ \mu_i^* \ge 0 & (4) \\ \mu_i^* g_i(\boldsymbol{x}^*) = 0 & (5) \end{cases}$$

参考文献:王燕军，梁治安. 最优化基础理论与方法[M]. 复旦大学出版社，2011.

https://en.wikipedia.org/wiki/Karush–Kuhn–Tucker_conditions

SVM对偶问题

Dual problem of SVM

$$
\begin{aligned}
\min_{\boldsymbol{w},b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\
\text{s.t.} \quad & 1 - y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \leqslant 0, \quad i = 1,2,\ldots,m
\end{aligned}
$$

SVM的KKT条件（充分+必要，充分性的证明参见参考文献中的定理4.4.4）：

$$
\begin{cases}
\nabla_{\boldsymbol{w},b}L(\boldsymbol{w}^*, b^*, \boldsymbol{\alpha}^*) = \nabla_{\boldsymbol{w},b}\left(\frac{1}{2}\|\boldsymbol{w}^*\|^2 + \sum_{i=1}^{m}\alpha_i(1 - y_i({\boldsymbol{w}^*}^T\boldsymbol{x}_i + b^*))\right) = 0 & (1) \\
1 - y_i({\boldsymbol{w}^*}^T\boldsymbol{x}_i + b^*) \leq 0 & (3) \\
\alpha_i^* \geq 0 & (4) \\
\alpha_i^*(1 - y_i({\boldsymbol{w}^*}^T\boldsymbol{x}_i + b^*)) = 0 & (5)
\end{cases}
$$

此即为式6.13

参考文献:王燕军，梁治安. 最优化基础理论与方法[M]. 复旦大学出版社，2011.

结 语

在这次课程中，我们学习了西瓜书

支持向量机的公式推导

那么在下次课程中，我们将会学习西瓜书

软间隔与支持向量回归

deepshare.net
深度之眼

联系我们:

电话: 18001992849

邮箱: service@deepshare.net

Q Q: 2677693114

公众号

客服微信