SVM 的数学推导和 Python 实现

赵新锋

2021 年 7 月 12 日

摘要

支持向量机（support vector machines, SVM）是一种分类模型，该模型在特征空间中求解间隔最大的分类超平面。

关键词： 支持向量机; SVM; SMO; 矩阵运算; 矩阵求导;numpy;sklearn.

目录

1 数学推导与 python 实现 **1**
1.1 线性可分 1
1.2 公式推导 1

1 数学推导与 python 实现

1.1 线性可分

简单说明如下：

1.2 公式推导

过程表示如下：

$$\begin{aligned} \arg\max_{w_0}\ d = \frac{y_0 \cdot (x_0^T \cdot w_0 + b_0)}{\|w_0\|} \\ \text{s.t.}\ \ \frac{y_i \cdot (x_i^T \cdot w_0 + b_0)}{\|w_0\|} \geq d \end{aligned} \tag{1}$$

将 w_0 和 b_0 进行一定比例的缩放

$$\begin{aligned} w &= \frac{w_0}{y_0 \cdot (x_0^T \cdot w_0 + b_0)} \\ b &= \frac{b_0}{y_0 \cdot (x_0^T \cdot w_0 + b_0)} \end{aligned}$$

可以将 (1) 式化简为：

$$\begin{aligned} &\arg\max_{w}\ d = \frac{1}{\|w\|} \\ \Longleftrightarrow\ &\arg\min_{w}\ d = \|w\| \\ \Longleftrightarrow\ &\arg\min_{w}\ d = \frac{1}{2} w^T \cdot w \qquad (2) \\ &\text{s.t.}\ \ y_i \cdot (x_i^T \cdot w + b) \geq 1 \qquad (3) \end{aligned}$$

将(2) 和 (3) 利用拉格朗日乘数法，获得拉格朗日原始问题形式：

$$\begin{aligned} \arg\min_{w,b}\max_{\alpha}\ L(w,b,\alpha) &= \frac{1}{2} w^T \cdot w - \alpha^T \cdot (y \odot (X \cdot w + b) - 1) \\ &= \frac{1}{2} w^T \cdot w - (\alpha \odot y)^T \cdot (X \cdot w + b) - \alpha^T \cdot \mathbf{1}^m \\ &= \frac{1}{2} w^T \cdot w - (\alpha \odot y)^T \cdot (X \cdot w + b) - \mathbf{1}^T \cdot \alpha \qquad (4) \end{aligned}$$

当满足 KTT 条件时，(4) 等价于拉格朗日对偶问题:

$$\arg\max_{\alpha}\min_{w,b}\ L(w,b,\alpha) = \frac{1}{2} w^T \cdot w - (\alpha \odot y)^T \cdot (X \cdot w + b) - \mathbf{1}^T \cdot \alpha \tag{5}$$

首先求 L 以 w、b 为参数的极小值，通过求微分得到偏导数形式。

$$\begin{aligned} \mathrm{d}L = &\frac{1}{2} tr[(\mathrm{d}w)^T \cdot w + w^T \cdot \mathrm{d}w) - (\alpha \odot y)^T \cdot (X \mathrm{d}w) \\ &- (\alpha \odot y)^T \cdot \mathrm{d}b \\ &- (\mathrm{d}\alpha)^T \cdot (y \odot (X \cdot w + b) - 1)] \\ = &\, tr[w^T \mathrm{d}w - (X^T \cdot (\alpha \odot y))^T \mathrm{d}w - (\alpha \odot y)^T \cdot \mathrm{d}b - (y \odot (X \cdot w + b) - 1)^T \mathrm{d}\alpha] \end{aligned}$$

从而得到 w、b 偏导数，并令偏导数为 0。

$$\frac{\partial L}{\partial w} = w - X^T \cdot (\alpha \odot y) = \mathbf{0}$$
$$\frac{\partial L}{\partial b} = -\alpha \odot y = \mathbf{0} \tag{6}$$

推导出如下等式和约束：

$$w = X^T \cdot (\alpha \odot y) \tag{7}$$

$$\alpha \odot y = \mathbf{0} \tag{8}$$

将(7)式和 (8)式代入(5)式中，

$$\begin{aligned}\arg\max_{\alpha} L(w, b, \alpha) &= \frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\ &\quad - (\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\ &\quad - (\alpha \odot y)^T \cdot b^m \\ &= -\frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) + (\alpha \odot y^T) \cdot b \\ &= -\frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y)\end{aligned}$$

去除符号，将极大转换成极小形式：

$$\arg\min_{\alpha} L(w, b, \alpha) = \frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \tag{9}$$
$$\text{s.t.} \quad \alpha \odot y = \mathbf{0}$$

加入软间隔参数, 每个向量点距离分类超平面距离增加 ξ_i，同时增加一个惩罚系数 C, 代入(5)新形式如下：

$$L(w, b, \xi, \alpha, u) = \frac{1}{2}w^T \cdot w + C \cdot \xi^T \cdot \mathbf{1}^m - \alpha^T \cdot (y \odot (X \cdot w + b) - 1) - u^T \cdot \xi \tag{10}$$

求得对于 w、b、ξ 的偏导并令其为 0：

$$\frac{\partial L}{\partial w} = w - X^T \cdot (\alpha \odot y) = \mathbf{0}$$
$$\frac{\partial L}{\partial b} = -\alpha \odot y = \mathbf{0}$$
$$\frac{\partial L}{\partial \xi} = C - \alpha - u = \mathbf{0}$$

代入 (10) 式中：

$$\arg\max_{\alpha} L(w, b, \xi, \alpha, u) = -\frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) + \mathbf{1}^m \cdot \alpha \tag{11}$$
$$\begin{aligned}\text{s.t.} \quad & \alpha \odot y = \mathbf{0} \\ & C - \alpha - u = \mathbf{0} \\ & \alpha_i \geq 0 \\ & u_i \geq 0\end{aligned}$$

End!!