

SVM 的数学推导和 Python 实现

赵新锋

2021 年 7 月 14 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。当训练数据近似线性可分时, 可以通过增加软间隔学习一个线性分类器。当线性不可分时, 利用核技巧, 隐式的将特征空间映射到高维特征空间, 从而达到线性可分。使用序列最小最优化算法 (SMO), 可以快速求解模型的参数。

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 分类超平面

当训练数据线性可分，可以得到一个线性超平面 $x \cdot w + b = 0$ ，将在超平面上方的归为正类，将在超平面下方的归为负类。当数据点与超平面距离越远时，表示分类的确定性越高，这样虽然线性分类的超平面可能有无数多个，但是我们可以找到一个所有点距离超平面最大的一个超平面。相应的决策函数为：

$$f(x) = \text{sign}(x \cdot w^* + b^*)$$

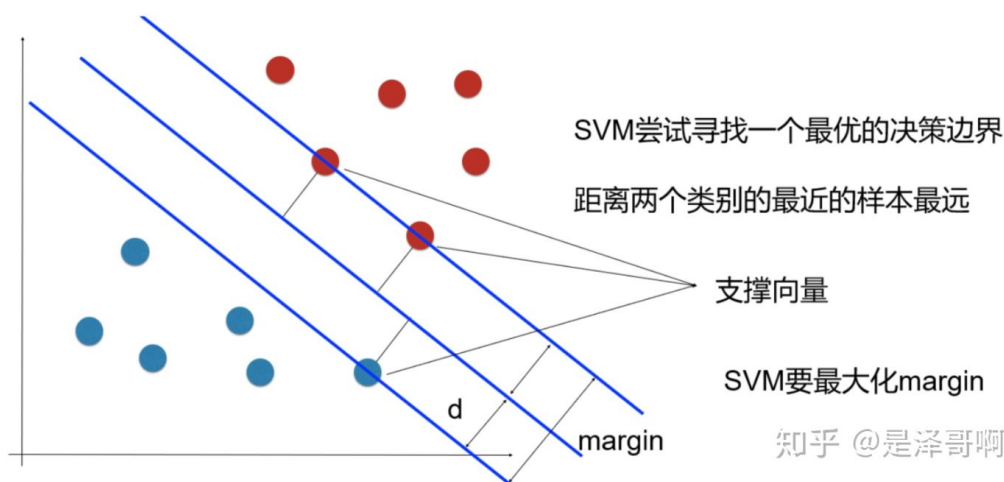


图 1: SVM 线性可分

1.2 公式推导

$x_i^T \cdot w_0 + b_0$ 为正时, y_i 为正, 当 $x_i^T \cdot w_0 + b_0$ 为负时, y_i 为负, 则可以定义 $\frac{y_i \cdot (x_i^T \cdot w_0 + b_0)}{\|w_0\|}$ 为几何间隔, 表示数据点距离超平面的距离。模型最终会找到一个参数为 w_0 和 b_0 的分离超平面, 所有点距离超平面的距离都大于等于 d , 将距离正好等于 d 的数据点称之为支持向量。

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

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 (??) 式化简为:

$$\begin{aligned} \arg \max_w d &= \frac{1}{\|w\|} \\ \iff \arg \min_w d &= \|w\| \\ \iff \arg \min_w d &= \frac{1}{2} w^T \cdot w \end{aligned} \quad (2)$$

$$\text{s.t. } y_i \cdot (x_i^T \cdot w + b) \geq 1 \quad (3)$$

将(??) 和 (??) 利用拉格朗日乘数法, 获得拉格朗日原始问题形式:

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

当满足 KKT 条件时, 拉格朗日对偶问题的解等价于(??) 的解:

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

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。

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

从而得到 w、b 偏导数, 并令偏导数为 0。

$$\begin{aligned} \frac{\partial L}{\partial w} &= w - X^T \cdot (\alpha \odot y) = \mathbf{0} \\ \frac{\partial L}{\partial b} &= -1^T \cdot (\alpha \odot y) = -y^T \cdot \alpha = 0 \end{aligned}$$

推导出如下关系:

$$w = X^T \cdot (\alpha \odot y) \quad (6)$$

$$y^T \cdot \alpha = 0 \quad (7)$$

将(??)式和 (??)式代入(??)式中,

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

去除负号，将极大转换成极小形式：

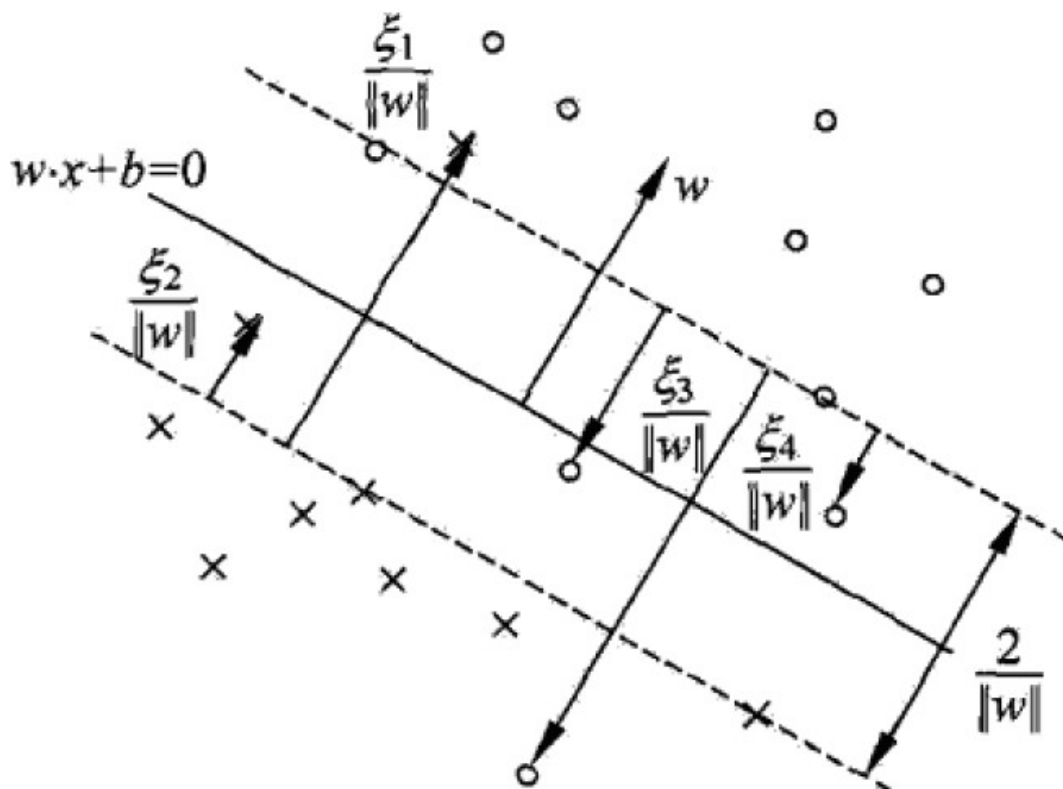
$$\begin{aligned} \arg \min_{\alpha} L(w, b, \alpha) &= \frac{1}{2} (\alpha \odot y)^T \cdot X \cdot X^T \cdot (\alpha \odot y) - \mathbf{1}^T \cdot \alpha \\ \text{s.t. } y^T \cdot \alpha &= 0 \end{aligned} \quad (8)$$

为了使拉格朗日对偶问题的解与原始问题解相同，需要同时满足 KKT 条件：

$$\begin{aligned} \frac{\partial L}{\partial w} &= 0 \\ \frac{\partial L}{\partial b} &= 0 \\ \alpha \odot (y \odot (X \cdot w + b) - 1) &= \mathbf{0}^m \\ y \odot ((X \cdot w + b) - 1) &\geq \mathbf{0}^m \\ \alpha &\geq \mathbf{0}^m \end{aligned}$$

1.3 软间隔

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(??)新形式如下：



软间隔的支持向量

图 2: SVM 软间隔

$$L(w, b, \xi, \alpha, \mu) = \frac{1}{2}w^T \cdot w + C \cdot \mathbf{1}^T \cdot \xi - \alpha^T \cdot (y \odot (X \cdot w + b) - 1) - \mu^T \cdot \xi \quad (9)$$

求得对于 w 、 b 、 ξ 的偏导并令其为 0:

$$\begin{aligned} \frac{\partial L}{\partial w} &= w - X^T \cdot (\alpha \odot y) = \mathbf{0} \\ \frac{\partial L}{\partial b} &= -\mathbf{1}^T \cdot (\alpha \odot y) = -y^T \cdot \alpha = 0 \\ \frac{\partial L}{\partial \xi} &= C - \alpha - \mu = \mathbf{0} \end{aligned}$$

代入 (??) 式中:

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

$C - \alpha - \mu = \mathbf{0}$ 、 $\alpha_i \geq 0$ 、 $\mu_i \geq 0$ 约束可以化简为:

$$\begin{aligned} \arg \max_{\alpha} L(w, b, \xi, \alpha, \mu) &= -\frac{1}{2}(\alpha \odot y)^T \cdot X \cdot X^T \cdot (\alpha \odot y) + \mathbf{1}^m \cdot \alpha \\ \text{s.t. } y^T \cdot \alpha &= 0 \\ 0 \leq \alpha_i &\leq C \end{aligned} \quad (10)$$

为了使拉格朗日对偶问题的解与原始问题解相同，需要同时满足 KKT 条件:

$$\begin{aligned} \frac{\partial L}{\partial w} &= 0 \\ \frac{\partial L}{\partial b} &= 0 \\ \frac{\partial L}{\partial \xi} &= 0 \\ \alpha \odot (y \odot (X \cdot w + b) - 1 + \xi) &= \mathbf{0}^m \\ y \odot ((X \cdot w + b) - 1 + \xi) &\geq \mathbf{0}^m \\ \alpha &\geq \mathbf{0}^m \\ \mu \odot \xi &= \mathbf{0}^m \\ \xi &\geq \mathbf{0}^m \\ \mu &\geq \mathbf{0}^m \end{aligned}$$

End!!