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 线性可分

三层网络简单说明如下：

- 输入层，不做处理，只是输入数据
- 隐藏层，单元数为 n_h 我们就以 RELU 为激活函数，另外一个常用的是 sigmoid，我们把激活函数封装成一个类，可以随时替换
- 输出层为一个单元，由于是二分类，我们就用 sigmoid 拟合对应类 1 的概率

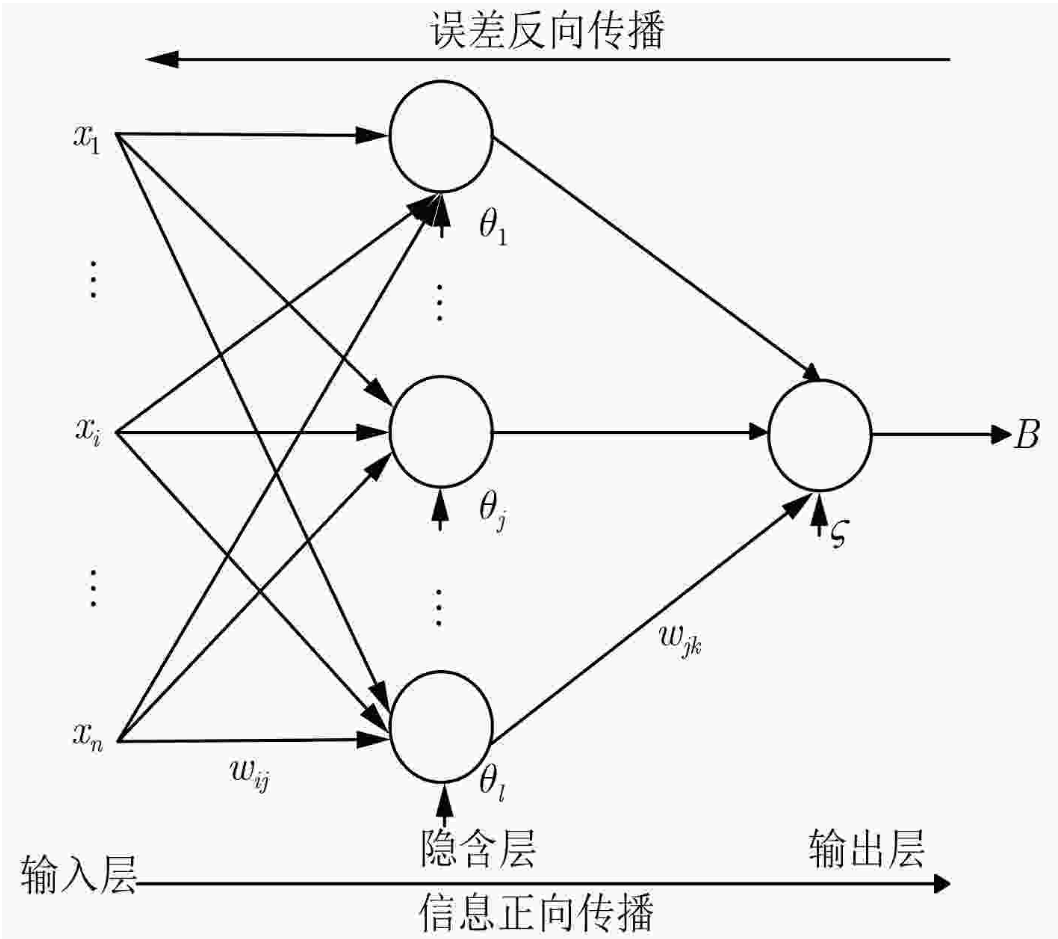

图 1: 神经网络示意图

- X 表示输入样本向量矩阵，Y 表示样本集的期望输出标记（有可能是 0/1 标记的向量，有可能是 onehot 的矩阵），本文示例由于是二分类，那么 Y 的每个元素要么是 0 要么是 1，Y 是 m 个元素的向量，但是在运算中为了通用将其转换为 $m \times 1$ 的矩阵。
- m 为样本数量,n 为样本向量的维度。

- 激活函数用 f 表示，隐藏层激活函数 f_h 使用 RELU 函数:$f_h(x) = max(0, x)$，输出层 f_o 为 sigmoid 函数，输出为 1 类的概率 $f_o(x) = \frac{1}{1+exp(-x)}$
- 隐藏层输出为 Y_h, 输出层的输出为 Y_o
- 隐藏层单元的参数为 w_h, N_h 个隐藏层单元的参数矩阵为 W_h, 输出层单元的参数为 w_o，N_o 个输出层单元的参数矩阵为 W_o。
- 虽然本文使用二分类即输出单元为 1 个，但是仍然使用矩阵 Yo 和 W_o 表示，从而使得算法是通用的，改成 onehot+softmax 方式的多分类，也是适用的。
- 从输入到输出的公式：$Y_o = f_o(f_h(X \cdot W_h) \cdot Wo)$
- 注：这里的线性变换 $X \cdot W_h$ 并没有使用偏置，通过在 X 增加一个全 1 的列，达到同样的效果。

1.2 公式推导

单个样本的运算过程表示如下:

$$
\begin{aligned}
\arg\max_{w_0}\ d &= \frac{y_0 \cdot (x_0^T \cdot w_0 + b_0)}{\|w_0\|} \\
\text{s.t.}\ \ \frac{y_i \cdot (x_i^T \cdot w_0 + b_0)}{\|w_0\|} &\geq d \\
w &= \frac{w_0}{y_0 \cdot (x_0^T \cdot w_0 + b_0)} \\
b &= \frac{b_0}{y_0 \cdot (x_0^T \cdot w_0 + b_0)} \\
\arg\max_{w}\ d &= \frac{1}{\|w\|} \iff \arg\min_{w}\ \|w\| \\
\text{s.t.}\ \ y_i \cdot (x_i^T \cdot w + b) &\geq 1 \\
\arg\min_{w,b}\ L(w,b,\alpha) &= \frac{1}{2} w^T \cdot w - \alpha^T \cdot (y \odot (X \cdot w + b) - 1) \\
&= \frac{1}{2} w^T \cdot w - (\alpha \odot y)^T \cdot (X \cdot w + b) - \alpha^T \cdot \mathbf{1}^m \\
&= \frac{1}{2} w^T \cdot w - (\alpha \odot y)^T \cdot (X \cdot w + b) - \mathbf{1}^T \cdot \alpha \\
\mathrm{d}L &= \frac{1}{2}((\mathrm{d}w)^T \cdot w + w^T \cdot \mathrm{d}w) - (\alpha \odot y)^T \cdot (X \mathrm{d}w) \\
&\quad - (\alpha \odot y)^T \cdot \mathrm{d}b \\
&\quad - (\mathrm{d}\alpha)^T \cdot (y \odot (X \cdot w + b) - 1) \\
&= w^T \mathrm{d}w - (X^T \cdot (\alpha \odot y))^T \mathrm{d}w - (\alpha \odot y)^T \cdot \mathrm{d}b - (y \odot (X \cdot w + b) - 1)^T \mathrm{d}\alpha \\
\frac{\partial L}{\partial w} &= w - X^T \cdot (\alpha \odot y) \\
\frac{\partial L}{\partial b} &= -\alpha \odot y \\
&\iff \\
w &= X^T \cdot (\alpha \odot y) \\
\alpha \odot y &= \mathbf{0} \\
\arg\max_{\alpha} L(w,b,\alpha) &= \frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\
&\quad - (\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\
&\quad - (\alpha \odot y)^T \cdot b^m \\
&= -\frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) + (\alpha \odot y^T) \cdot b \\
&= -\frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\
\iff \arg\min_{\alpha} L(w,b,\alpha) &= \frac{1}{2}(\alpha \odot y)^T \cdot X^T \cdot X \cdot (\alpha \odot y) \\
\text{s.t.}\ \ \alpha \odot y &= \mathbf{0}
\end{aligned} \tag{1}
$$

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

End!!