

支持向量机

support vector machine
algorithm verify

我真的不懂忧郁



支持向量机

support vector machine
algorithm verify

by

我真的不懂忧郁

Student Name	Student Number
First Surname	1234567

Instructor:	I. Surname
Teaching Assistant:	I. Surname
Project Duration:	Month, Year - Month, Year
Faculty:	Faculty of Aerospace Engineering, Delft

Cover: Canadarm 2 Robotic Arm Grapples SpaceX Dragon by NASA under
CC BY-NC 2.0 (Modified)

Style: TU Delft Report Style, with modifications by Daan Zwaneveld

Preface

A preface...

我真的不懂忧郁
Delft, April 2024

Summary

A summary...

目录

Preface	i
Summary	ii
Nomenclature	iv
1 拉格朗日对偶性	1
1.1 极小极大值问题（原始问题）	1
1.2 极大极小值问题（对偶问题）	2
1.3 原始问题和对偶问题的关系	2
1.4 KKT 条件	3
2 硬间隔最大化与线性可分向量机	4
2.1 线性可分支持向量机	4
2.2 函数间隔和几何间隔	4
2.3 最大间隔分离超平面	6
2.4 最大间隔超平面的存在性和唯一性	6
2.5 支持向量和间隔边界	6
2.6 支持向量机的对偶算法	7
3 软间隔最大化和线性支持向量机	9
4 核函数和非线性支持向量机	10
References	11
A Source Code Example	12
B Task Division Example	13

Nomenclature

If a nomenclature is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

Abbreviations

Abbreviation	Definition
ISA	International Standard Atmosphere
...	

Symbols

Symbol	Definition	Unit
V	Velocity	[m/s]
...		
ρ	Density	[kg/m ³]
...		

Chapter 1

拉格朗日对偶性

1.1. 极小极大值问题（原始问题）

假设 $f(x), c_i(x), h_j(x)$ 是定义在 \mathbb{R}^n 上的连续可微函数。考虑优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0, \quad i = 1, 2, \dots, k; \\ & h_j(x) = 0, \quad j = 1, 2, \dots, l; \end{aligned} \quad (1.1)$$

引入拉格朗日函数

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^k \alpha_i c_i(x) + \sum_{j=1}^l \beta_j h_j(x) \quad (1.2)$$

其中 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T \in \mathbb{R}^n$, α_i, β_j 是拉格朗日乘子, $\alpha_i \geq 0$ 。考虑

$$\theta_P(x) = \max_{\alpha, \beta, \alpha_i \geq 0} L(x, \alpha, \beta) \quad (1.3)$$

其中下标 P 表示原始问题。

假设给定某个 x , 如果 x 违反原始问题的约束条件, 即存在 i 使得 $c_i(x) > 0$ 或存在 j 使得 $h_j(x) \neq 0$, 则

$$\theta_P(x) = \max_{\alpha, \beta, \alpha_i \geq 0} [f(x) + \sum_{i=1}^k \alpha_i c_i + \sum_{j=1}^l \beta_j h_j(x)] = +\infty \quad (1.4)$$

因为若 $c_i(x) > 0$, 那么令 $\alpha_i \rightarrow +\infty$, 若 $h_j(x) \neq 0$, 则令 β_j 使得 $\beta_j h_j(x) \rightarrow +\infty$, 其余将 α_i, β_i 取为 0 即可。

因此

$$\theta_P(x) = \begin{cases} f(x), & x \text{ 满足原始问题约束} \\ +\infty, & \text{else} \end{cases} \quad (1.5)$$

所以如果考虑极小化问题

$$\theta_P(x) = \min_x \max_{\alpha, \beta, \alpha_i \geq 0} L(x, \alpha, \beta) \quad (1.6)$$

它与原始优化问题是等价的，即解相同。问题 $\min_x \max_{\alpha, \beta, \alpha_i \geq 0} L(x, \alpha, \beta)$ 也被称为广义拉格朗日函数的极小极大问题。为了方便，定义原始问题为最优值

$$p^* = \min_x \theta_P(x) \quad (1.7)$$

为原始问题的值。

1.2. 极大极小值问题（对偶问题）

定义

$$\theta_D(\alpha, \beta) = \min_x L(x, \alpha, \beta) \quad (1.8)$$

在考虑极大化 $\theta_D(\alpha, \beta)$

$$\max_{\alpha, \beta, \alpha_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha, \beta, \alpha_i \geq 0} \min_x L(x, \alpha, \beta) \quad (1.9)$$

这样的问题称为广义拉格朗日函数的极大极小问题

可以将广义拉格朗日函数的极大极小问题表示为约束最优化问题：

$$\begin{aligned} \max_{\alpha, \beta, \alpha_i \geq 0} \theta_D(\alpha, \beta) &= \max_{\alpha, \beta, \alpha_i \geq 0} \min_x L(x, \alpha, \beta) \\ \text{s.t. } \alpha_i &\geq 0, \quad i = 1, 2, \dots, k \end{aligned} \quad (1.10)$$

称为原始问题的对偶问题，定义对偶问题的最优值

$$d^* = \max_{\alpha, \beta, \alpha_i \geq 0} \theta_D(\alpha, \beta) \quad (1.11)$$

称为对偶问题的值。

1.3. 原始问题和对偶问题的关系

theorem 1.3.1: 若原始问题和对偶问题都有最优值，则

$$d^* = \max_{\alpha, \beta, \alpha_i \geq 0} \theta_D(\alpha, \beta) \leq \min_{\alpha, \beta, \alpha_i \geq 0} \theta_P(\alpha, \beta) = p^* \quad (1.12)$$

corollary 1.3.2: 设 x^* 和 α^*, β^* 分别是原始问题和对偶问题的可行解，并且 $d^* = p^*$ ，则 x^* 和 α^*, β^* 分别是原始问题和对偶问题的最优解。

theorem 1.3.3: 考虑原始问题和对偶问题。假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数， $h_j(x)$ 是仿射函数；并且假设不等式约束 $c_i(x)$ 是严格可行的，即存在 x_i 对所有的 i 有 $c_i(x) < 0$ ，则存在 x^*, α^*, β^* ，使得 x^* 是原始问题的解， α^*, β^* 是对偶问题的解，并且

$$p^* = d^* = L(x^*, \alpha^*, \beta^*) \quad (1.13)$$

1.4. KKT 条件

theorem 1.4.1: 对原始问题和对偶问题。假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数, $h_j(x)$ 是仿射函数; 并且不等式约束 $c_i(x)$ 是严格可行的, 则 x^* 和 α^*, β^* 分别是原始问题和对偶问题的解的充分必要条件是: α^*, β^*, x^* 满足下面的 **Karush-Kuhn-Tucker** 条件 (KKT)

$$\nabla_x L(x^*, \alpha^*, \beta^*) = 0 \quad (1.14)$$

$$\alpha_i^* c_i(x^*) = 0, \quad i = 1, 2, \dots, k \quad (1.15)$$

$$c_i(x^*) \leq 0, \quad i = 1, 2, \dots, k \quad (1.16)$$

$$\alpha_i^* \geq 0, \quad i = 1, 2, \dots, k \quad (1.17)$$

$$h_j(x^*) = 0, \quad i = 1, 2, \dots, l \quad (1.18)$$

其中 (1.15) 是 KKT 的对偶互补条件。由此条件, 若 $\alpha_i^* > 0$, 则 $c_i(x^*) = 0$

Chapter 2

硬间隔最大化与线性可分向量机

Keyword: 硬间隔最大化、软间隔最大化、核技巧

2.1. 线性可分支持向量机

definition 2.1.1: (线性可分支持向量机) 给定线性可分训练数据集, 通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为

$$\omega^* \cdot x + b^* = 0 \quad (2.1)$$

以及相应的分类决策函数

$$f(x) = \text{sign}(\omega^* \cdot x + b^*) \quad (2.2)$$

称为线性可分支持向量机。

线性支持向量机假设输入空间和特征空间的元素一一对应, 并将输入空间中的输入映射为特征空间中的特征向量。非线性支持向量机利用一个从输入空间到特征空间的非线性映射将输入映射为特征向量。所以支持向量机的学习都是在特征空间中进行的。

一般地, 当训练数据集线性可分时, 存在无穷个分离超平面可以将两类数据正确分开。感知机利用误分类最小策略, 求得分离超平面, 不过这个时候的解有无数个。线性可分支持向量机利用间隔最大化求最有分离超平面, 这个时候的解是唯一的。

2.2. 函数间隔和几何间隔

函数间隔

definition 2.2.1: (函数间隔) 对于给定的训练数据集 T 和超平面 (ω, b) , 定义超平面 (ω, b) 关于样本点 (x_i, y_i) 的函数间隔为

$$\hat{\gamma}_i = y_i(\omega \cdot x_i + b) \quad (2.3)$$

定义超平面 (ω, b) 关于训练数据集 T 的函数间隔为超平面 (ω, b) 关于 T 中所有样本点的函数间隔的最小值

$$\hat{\gamma}_i = \min_{j=1, \dots, N} \hat{\gamma}_j \quad (2.4)$$

函数间隔可以表示分类预测的正确性和确信程度。但是选择分离超平面时，我们发现只要等比例 ω 和 b ，分离超平面是不变的如：

$$\omega \cdot x + b = 0 \Leftrightarrow 2\omega \cdot x + 2b = 0 \quad (2.5)$$

但是此时函数间隔却变成了 2 倍

$$\hat{\gamma}_i = y_i(2\omega \cdot x_i + 2b) = 2y_i(\omega \cdot x_i + b) \quad (2.6)$$

几何间隔

因此，需要对分离超平面的法向量加某些约束，如规范化 $\|\omega\| = 1$ ，使得间隔是确定的，这时函数间隔成为几何间隔。

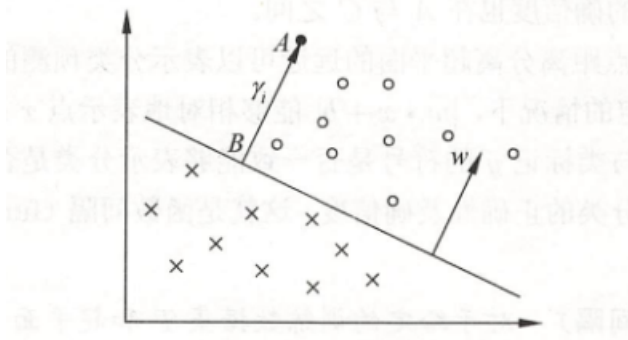


图 2.1: 几何间隔

给定超平面 (ω, b) 以及法向量 ω 。点 A 表示某一实例 x_i ，分类标记为 $y_i = \pm 1$ 。点 A 与超平面 (ω, b) 的距离由线段 AB 给出，记为 γ_i 。

$$\gamma_i = \pm \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \quad (2.7)$$

一般地，当样本点 (x_i, y_i) 被正确分类时，点 x_i 与超平面 (ω, b) 的距离为

$$\gamma_i = y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \quad (2.8)$$

definition 2.2.2: (几何间隔) 对于给定的训练数据集 T 和超平面 (ω, b) ，定义超平面 (ω, b) 关于样本点 (x_i, y_i) 的几何间隔为

$$\gamma_i = y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \quad (2.9)$$

定义超平面 (ω, b) 关于训练数据集 T 的几何间隔为超平面 (ω, b) 关于 T 中所有样本点的几何间隔的最小值

$$\gamma_i = \min_{j=1, \dots, N} \hat{\gamma}_j \quad (2.10)$$

函数间隔和几何间隔的关系

从函数间隔和几何间隔的定义，二者有如下关系

$$\gamma = \frac{\hat{\gamma}}{\|\omega\|} \quad (2.11)$$

可见如果 $\|\omega\| = 1$ ，则二者相等。如果超平面参数 ω 和 b 成比例改变，函数间隔也成比例改变而几何间隔不变。

2.3. 最大间隔分离超平面

如何求得几何间隔最大的分离超平面？这个问题表述为下列优化问题

$$\begin{aligned} \max_{\omega, b} \quad & \gamma \\ \text{s.t.} \quad & y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \geq \gamma_i, \quad i = 1, 2, \dots, N \end{aligned} \quad (2.12)$$

考虑几何间隔和函数间隔的关系，优化问题等价于

$$\begin{aligned} \max_{\omega, b} \quad & \frac{\hat{\gamma}}{\|\omega\|} \\ \text{s.t.} \quad & y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \geq \gamma_i, \quad i = 1, 2, \dots, N \end{aligned} \quad (2.13)$$

函数间隔 $\hat{\gamma}$ 的取值并不影响优化问题的解。事实上，假设 ω 和 b 成比例改变为 $\lambda\omega$ 和 λb ，这时函数间隔成为 $\lambda\hat{\gamma}$ 。函数间隔的这一改变对上面最优化问题的不等式约束没有影响，对目标函数的优化也没有影响，即它产生一个等价的最优化问题，这样就可以取 $\hat{\gamma} = 1$ 代入优化问题。注意到最大化 $1/\|\omega\|$ 和最小化 $\frac{1}{2}\|\omega\|^2$ 是等价的。于是就得到下面的线性可分支持向量机学习的最优化问题：

$$\begin{aligned} \max_{\omega, b} \quad & \frac{1}{2}\|\omega\|^2 \\ \text{s.t.} \quad & y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \geq \gamma_i, \quad i = 1, 2, \dots, N \end{aligned} \quad (2.14)$$

这是一个凸二次规划问题。

2.4. 最大间隔超平面的存在性和唯一性

theorem 2.4.1: 若训练数据集 T 是线性可分的，则可将数据集中样本点完全正确分开的最大间隔分离超平面存在且唯一

2.5. 支持向量和间隔边界

训练数据集的样本点中与分离超平面距离最近的样本点的实例称为**支持向量** (support vector)。支持向量是使得约束条件 $y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right) \geq \gamma_i, \quad i = 1, 2, \dots, N$ 成立的点，即

$$y_i(\omega \cdot x_i + b) - 1 = 0 \quad (2.15)$$

对于 $y_1 = +1$ 的正例点，支持向量在超平面

$$H_1 : \omega \cdot x + b = 1 \quad (2.16)$$

对于 $y_1 = -1$ 的负例点，支持向量在超平面

$$H_2 : \omega \cdot x + b = -1 \quad (2.17)$$

如下图所示， H_1 和 H_2 上的点就是支持向量。

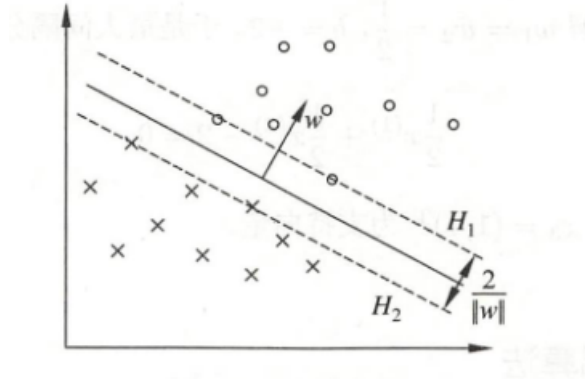


图 2.2: 支持向量

H_1 和 H_2 之间的距离称为间隔 (margin)。间隔依赖于分离超平面的法向量 ω ，等于 $2/\|\omega\|$ 。 H_1 和 H_2 称为间隔边界 (margin boundary)。

在决定分离超平面时，只有支持向量起作用，而其他实例点不起作用。所以支持向量机由很少的支持向量作为训练样本决定。

2.6. 支持向量机的对偶算法

应用拉格朗日对偶性，通过求解对偶问题 (dual problem) 得到原始问题的最优解。这就是线性可分支持向量机的对偶算法。

首先构建拉格朗日函数 (Lagrange function)。引入拉格朗日乘子 (Lagrange multiplier): $\alpha_1 \geq 0$ 。

$$L(\omega, b, \alpha_i) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^N \alpha_i y_i (\omega \cdot x_i + b) + \sum_{i=1}^N \alpha_i \quad (2.18)$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ 为拉格朗日乘子向量。根据拉格朗日乘子法，原始问题的对偶问题是极大极小问题

$$\max_{\alpha} \min_{\omega, b} L(\omega, b, \alpha) \quad (2.19)$$

theorem 2.6.1: 设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ 是最优化对偶问题 (??对偶最优化问题)，的解，则存在下标 j ，使得 $\alpha_j^* > 0$ ，并可按照下式求得原始最优化问题的解 ω^*, b^* 。

$$\omega^* = \sum_{i=1}^N \alpha_i^* y_i x_i \quad (2.20)$$

$$b^* = y_j - \sum_{i=1}^N \alpha_i^* y_i (x_i \cdot x_j) \quad (2.21)$$

由此定理，分离超平面可以写成

$$\sum_{i=1}^N \alpha_i^* y_i (x \cdot x_i) + b^* = 0 \quad (2.22)$$

分类决策函数可以写成

$$f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i^* y_i (x \cdot x_i) + b^*\right) \quad (2.23)$$

这样分类决策函数只依赖于输入 x 和训练样本输入的内积。 $f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i^* y_i (x \cdot x_i) + b^*\right)$ 线性可分支持向量机的对偶形式。

线性可分支持向量机学习算法

输入：线性可分训练集 $T = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ，其中 $x_i \in \mathbb{R}^n$ ， $y_i \in \{-1, +1\}$ 。输出：分离超平面和分类决策函数

1. 构造并求解约束最优化问题

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) = \sum_{i=1}^N \alpha_i \\ .s.t \quad & \sum_{i=1}^N \alpha_i y_i = 0 \\ & \alpha_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned} \quad (2.24)$$

求得最优解 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$

2. 计算

$$\omega^* = \sum_{i=0}^N \alpha_i^* y_i x_i \quad (2.25)$$

并选择 α^* 的一个正分量 $\alpha_j^* > 0$ 计算

$$b^* = y_i \sum_{i=1}^N \alpha_i^* y_i (x_i \cdot x_j) \quad (2.26)$$

3. 求得分离超平面和决策函数

$$\omega^* \cdot x + b^* = 0 \quad (2.27)$$

$$f(x) = \text{sign}(\omega^* \cdot x + b^*) \quad (2.28)$$

线性可分支持向量机中， ω^* 和 b^* 只依赖于训练数据中对应 $\alpha_i^* > 0$ 的样本点 (x_i, y_i) ，我们将训练数据中对应于 $\alpha_i^* > 0$ 的实例点 $x_I \in \mathbb{R}^n$ 称为支持向量。

definition 2.6.1 (支持向量): 考虑原始最优化问题以及对偶最优化问题，将训练数据集集中对应于 $\alpha_i^* > 0$ 的样本点 (x_i, y_i) 的实例 $x_i \in \mathbb{R}^n$ 称为支持向量。

Chapter 3

软间隔最大化和线性支持向量机

Chapter 4

核函数和非线性支持向量机

References

- [1] I. Surname, I. Surname, and I. Surname. “The Title of the Article”. In: *The Title of the Journal* 1.2 (2000), pp. 123–456.

Chapter A

Source Code Example

Adding source code to your report/thesis is supported with the package listings. An example can be found below. Files can be added using `\lstinputlisting[language=<language>]{<filename>}`.

```
1 """
2 ISA Calculator: import the function, specify the height and it will return a
3 list in the following format: [Temperature,Density,Pressure,Speed of Sound].
4 Note that there is no check to see if the maximum altitude is reached.
5 """
6
7 import math
8 g0 = 9.80665
9 R = 287.0
10 layer1 = [0, 288.15, 101325.0]
11 alt = [0,11000,20000,32000,47000,51000,71000,86000]
12 a = [-.0065,0,.0010,.0028,0,-.0028,-.0020]
13
14 def atmosphere(h):
15     for i in range(0,len(alt)-1):
16         if h >= alt[i]:
17             layer0 = layer1[:]
18             layer1[0] = min(h,alt[i+1])
19             if a[i] != 0:
20                 layer1[1] = layer0[1] + a[i]*(layer1[0]-layer0[0])
21                 layer1[2] = layer0[2] * (layer1[1]/layer0[1])**(-g0/(a[i]*R))
22             else:
23                 layer1[2] = layer0[2]*math.exp((-g0/(R*layer1[1]))*(layer1[0]-layer0[0]))
24     return [layer1[1],layer1[2]/(R*layer1[1]),layer1[2],math.sqrt(1.4*R*layer1[1])]
```

Chapter B

Task Division Example

If a task division is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

表 B.1: Distribution of the workload

Task	Student Name(s)
Summary	
Chapter 1 Introduction	
Chapter 2	
Chapter 3	
Chapter *	
Chapter * Conclusion	
Editors	
CAD and Figures	
Document Design and Layout	