

机器学习

Machine learning

第三章 线性分类

Linear Classifier

授课人：周晓飞
zhouxiaofei@iie.ac.cn
2020-10-22

第三章 线性分类

3.1 概述

3.2 基础知识

3.3 感知机

3.4 线性鉴别分析

3.5 logistic 模型

3.6 其它方法：最小距离分类、最小均方误差方法、线性 SVM

第三章 线性分类

3.1 概述

3.2 基础知识

3.3 感知机

3.4 线性鉴别分析

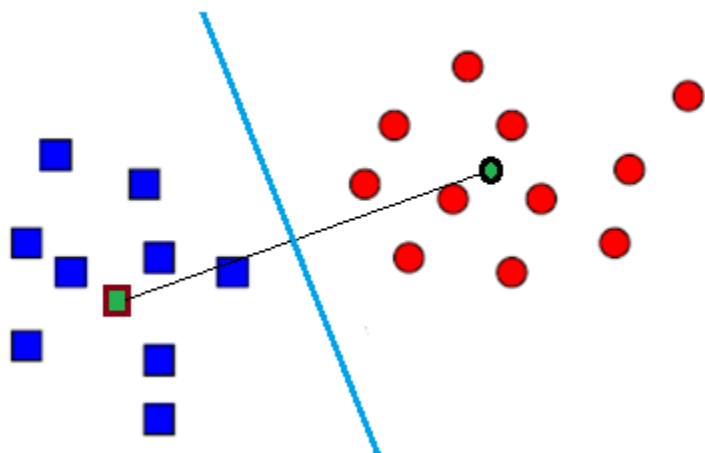
3.5 logistic 模型

3.6 其它方法：最小距离方法、最小均方误差方法、线性 SVM

最小距离分类

基本思想

每类建立一个原型向量，以最小距离的原型作为分类依据。
(文本分类 Rocchio 分类器).



最小距离分类

最小距离分类器

- 类别的均值向量作为原型向量:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{x}_k^{(i)}$$

- 决策准则: $h(\mathbf{x}) = \arg \min_{c_i} \|\mathbf{x} - \mathbf{m}_i\|^2$

最小距离分类

最小距离分类器

- 决策平面: $g(\mathbf{x}) = \mathbf{w}^T (\mathbf{x} - \mathbf{x}_0) = 0$,

$$\mathbf{w} = \mathbf{m}_1 - \mathbf{m}_2, \quad \mathbf{x}_0 = \frac{1}{2}(\mathbf{m}_1 + \mathbf{m}_2)$$

- 决策函数: $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$

$$\mathbf{w} = \mathbf{m}_1 - \mathbf{m}_2, \quad b = -\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T (\mathbf{m}_1 + \mathbf{m}_2)$$

第三章 线性分类

3.1 概述

3.2 基础知识

3.3 感知机

3.4 线性鉴别分析

3.5 logistic 模型

3.6 其它方法：最小距离方法、最小均方误差方法、线性 SVM

最小距离分类

基本思想

由线性模型拟合类别，最小化样本类别预测误差。
(等价于预测为离散值的线性回归模型)

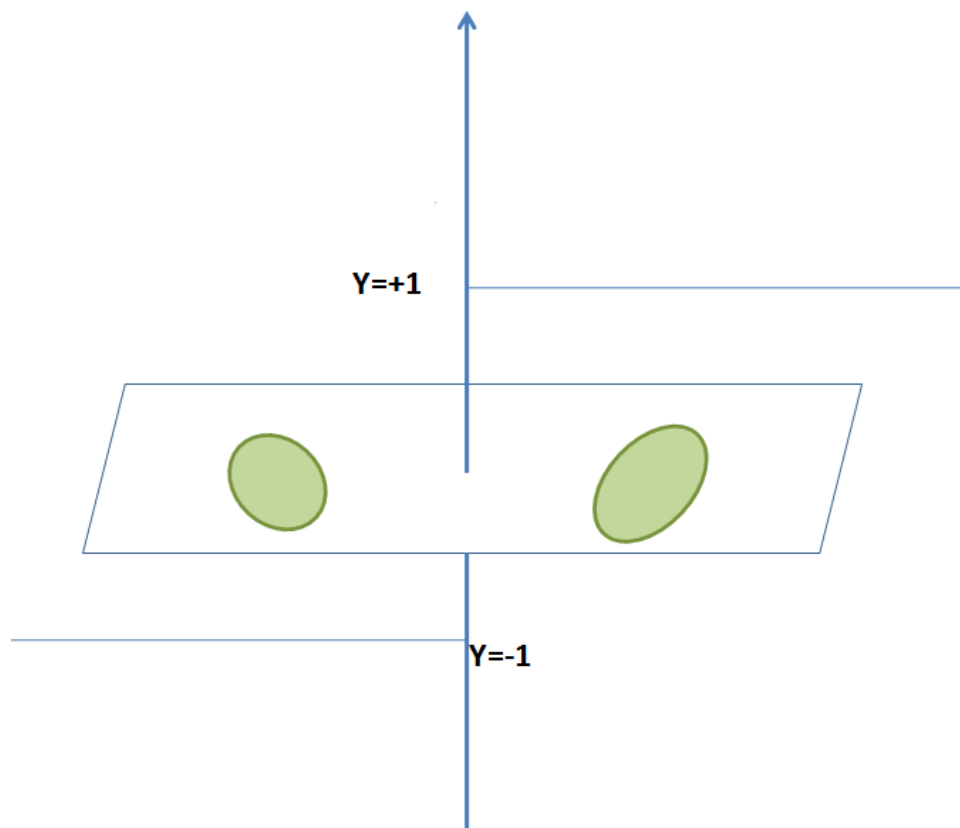
代价函数：

$$J(w) = \sum_{i=1}^N (y_i - \mathbf{x}_i^T w)^2 \equiv \sum_{i=1}^N e_i^2$$

其中， \mathbf{x} , w 为增广表示。

最小距离分类

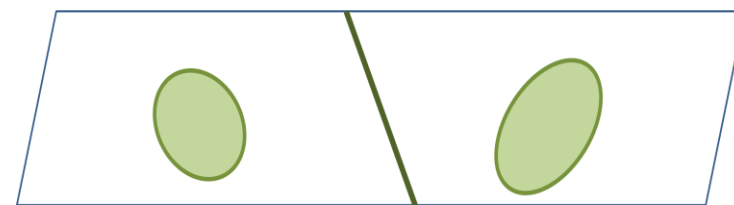
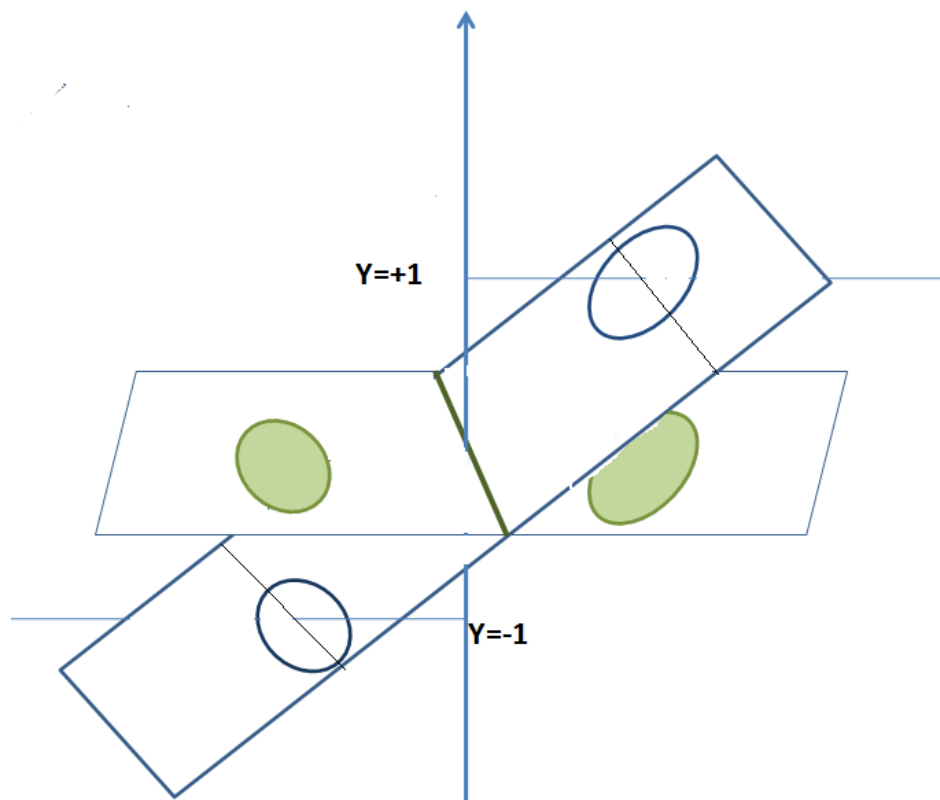
基本思想



思考：1. 从几何上看，决策界应当是什么样？
2. 类别标签 y 的取值对结果的影响？

最小距离分类

基本思想



- 思考：1. 从几何上看，决策界应当是什么样？
2. 类别标签 y 的取值对结果的影响？

最小距离分类

学习算法 1- 伪逆矩阵法

代价函数导数为 0:

$$\sum_{i=1}^N \mathbf{x}_i (y_i - \mathbf{x}_i^T \hat{\mathbf{w}}) = 0 \Rightarrow \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \right) \hat{\mathbf{w}} = \sum_{i=1}^N (\mathbf{x}_i y_i)$$

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1l} \\ x_{21} & x_{22} & \dots & x_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nl} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$(X^T X) \hat{\mathbf{w}} = X^T \mathbf{y} \Rightarrow \hat{\mathbf{w}} = (X^T X)^{-1} X^T \mathbf{y}$$

X 的伪逆 $X^\#$



最小距离分类

学习算法 1- 伪逆矩阵法

例子 1:

Class ω_1 consists of the two-dimensional vectors $[0.2, 0.7]^T$, $[0.3, 0.3]^T$, $[0.4, 0.5]^T$, $[0.6, 0.5]^T$, $[0.1, 0.4]^T$ and class ω_2 of $[0.4, 0.6]^T$, $[0.6, 0.2]^T$, $[0.7, 0.4]^T$, $[0.8, 0.6]^T$, $[0.7, 0.5]^T$. Design the sum of error squares optimal linear classifier $w_1x_1 + w_2x_2 + w_0 = 0$.

We first extend the given vectors by using 1 as their third dimension and form the 10×3 matrix X , which has as rows the transposes of these vectors. The resulting sample correlation 3×3 matrix $X^T X$ is equal to

$$X^T X = \begin{bmatrix} 2.8 & 2.24 & 4.8 \\ 2.24 & 2.41 & 4.7 \\ 4.8 & 4.7 & 10 \end{bmatrix}$$

最小距离分类

学习算法 1- 伪逆矩阵法

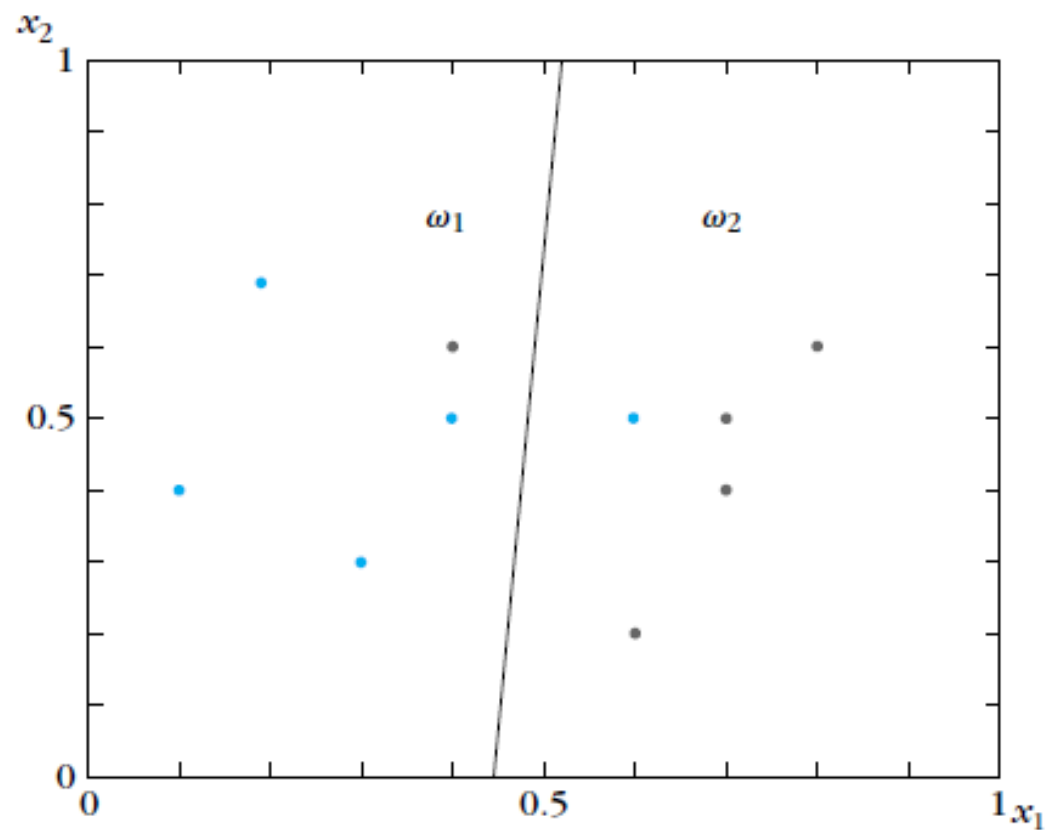
The corresponding \mathbf{y} consists of five 1's and then five -1's and

$$X^T \mathbf{y} = \begin{bmatrix} -1.6 \\ 0.1 \\ 0.0 \end{bmatrix}$$

$$\hat{\mathbf{w}} = (X^T X)^{-1} X^T \mathbf{y}$$

Solving the corresponding set of equations results in $[w_1, w_2, w_0] = [-3.218, 0.241, 1.431]$

最小距离分类



最小距离分类

学习算法 2-梯度下降算法

避免大规模矩阵求逆，采用梯度下降方法，称为最小均方(LMSE)方法。

$$\nabla = \left[\frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, \dots, \frac{\partial}{\partial w_n} \right]^T$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \eta(n) \nabla J(\mathbf{w}); \quad \eta(n) > 0$$

$$\nabla J(\mathbf{w}) = -e(n) \mathbf{x}(n)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n) e(n) \mathbf{x}(n)$$

第三章 线性分类

3.1 概述

3.2 基础知识

3.3 感知机

3.4 线性鉴别分析

3.5 logistic 模型

3.6 其它方法：最小距离方法、最小均方误差方法、**线性 SVM**

线性支持向量机

推荐书 1

支持向量机 (support vector machine, SVM) 是数据挖掘中的一项新技术, 是借助于最优化方法解决机器学习问题的新工具. 它最初于 20 世纪 90 年代由 Vapnik 提出, 近年来在其理论研究和算法实现方面都取得了突破性进展, 开始成为克服“维数灾难”和“过学习”等传统困难的有力手段. 虽然它还处于飞速发展的阶段, 但是它的理论基础和实现途径的基本框架已经形成. 自 2000 年开始, 国外已陆续有几本专著出版. 据我们所知, 本书是国内第一本专门对它进行全面完整介绍和论述的书籍.

数据挖掘中的新方法 ——支持向量机

邓乃扬 田英杰 著

推荐书 2:

Vladimir N. Vapnik

The Nature of Statistical Learning Theory

Second Edition

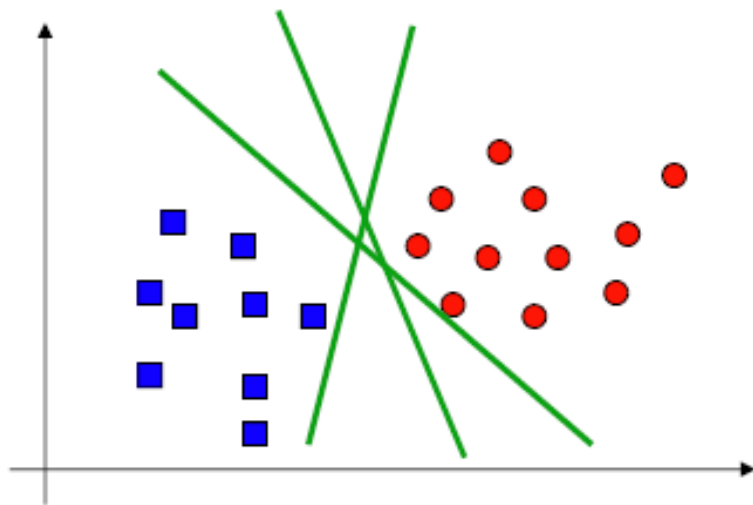
With 50 Illustrations

线性支持向量机

基本思想

找到最大间隔分类超平面。

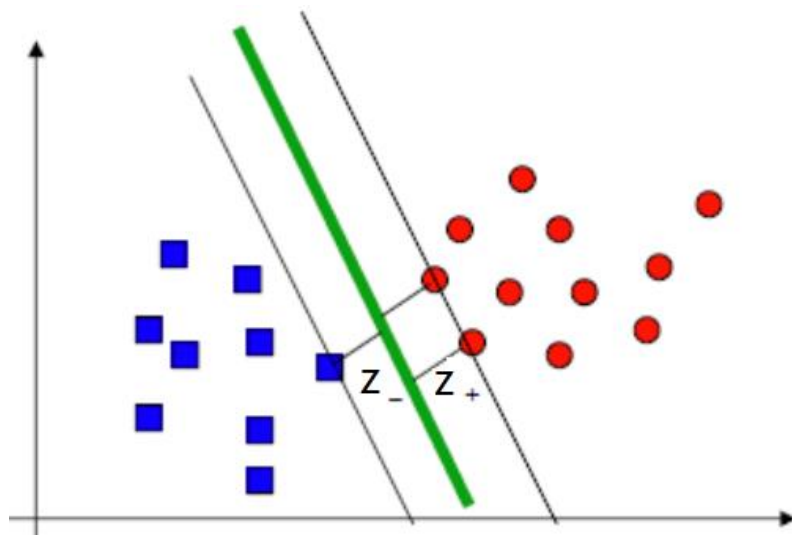
有多个超平面将数据分开，哪个更好？



线性支持向量机

基本思想

目标：类别集合到分类超平面的最小距离最大化

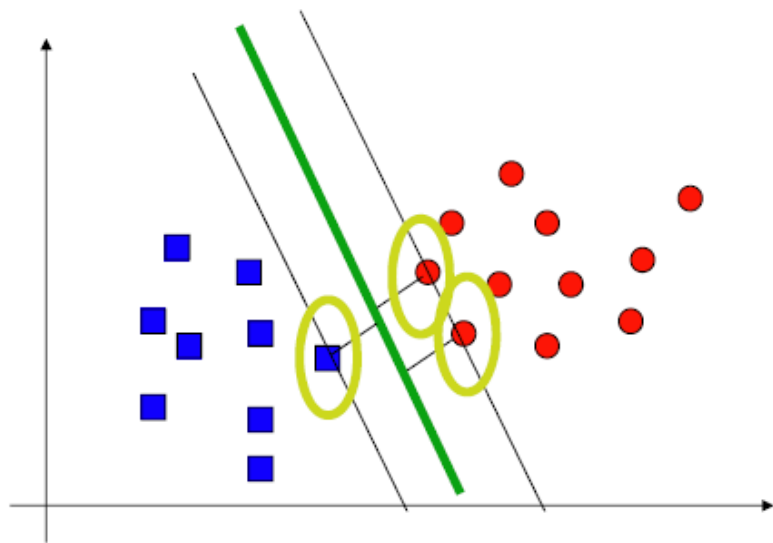


线性支持向量机

支持向量 (SV) 支撑超平面

支撑向量：支撑最小距离最大化的样本。

支撑超平面：由支持向量张成，平行于分类面的超平面。



线性支持向量机

分类间隔

样本到分类面的间隔

$$|z| = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|}$$

- $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$
- $|g(\mathbf{x})|$ 正比于 $\|\mathbf{w}\|$

线性支持向量机

分类间隔

SV 到分类超平面的间隔:

为了 SVM 优化目标有唯一解, 固定模长, 避免 w 方向重复, 固定 $|g(x)|$ 或固定 $|w|$ 。

选择固定 $|g(x)|$, 令 $|g(x)| = c = 1$ 则

$$g(x) = \pm 1$$

$$z = \frac{g(x)}{\|w\|}, \quad g(x) = \pm 1, \quad z_{\pm} = \frac{\pm 1}{\|w\|}$$

线性支持向量机

分类间隔

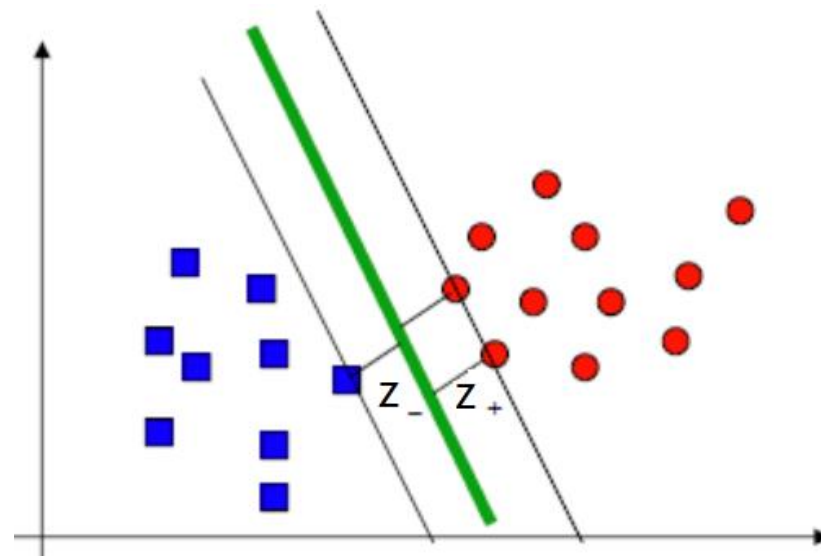
支撑超平面的表示:

$$\mathbf{w}^T \mathbf{x} + b = \pm 1$$

等价表达:

$$y(\mathbf{w}^T \mathbf{x} + b) - 1 = 0$$

y 为样本 \mathbf{x} 的类别标签。



线性支持向量机

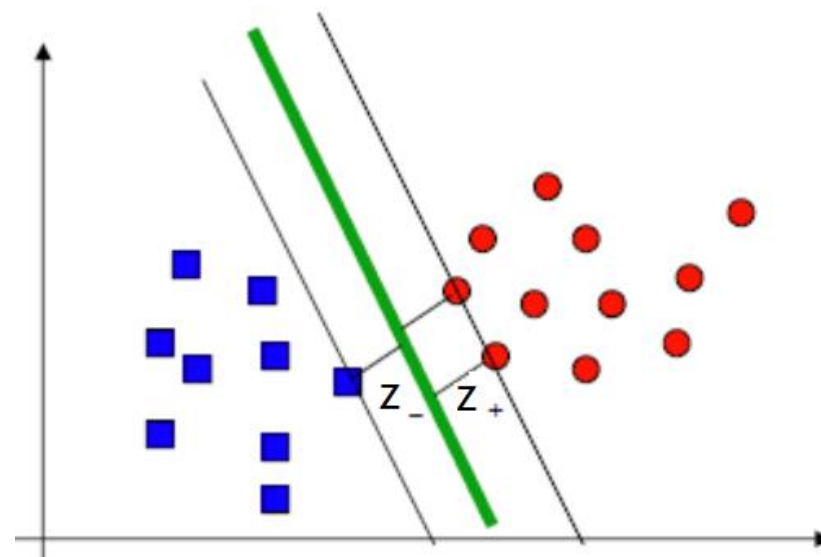
分类间隔

SVM 期望样本在支撑面的两侧：

For $(\mathbf{x}_i, y_i = +1)$, $\mathbf{w}^T \mathbf{x}_i + b \geq 1$, $\mathbf{x}_i \in \omega_+$;

For $(\mathbf{x}_i, y_i = -1)$, $\mathbf{w}^T \mathbf{x}_i + b \leq -1$, $\mathbf{x}_i \in \omega_-$;

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0$$



线性支持向量机

最大间隔

最大间隔问题：

$$\max_w d = \max(z_+ - z_-) = \max_w \frac{2}{\|w\|}$$

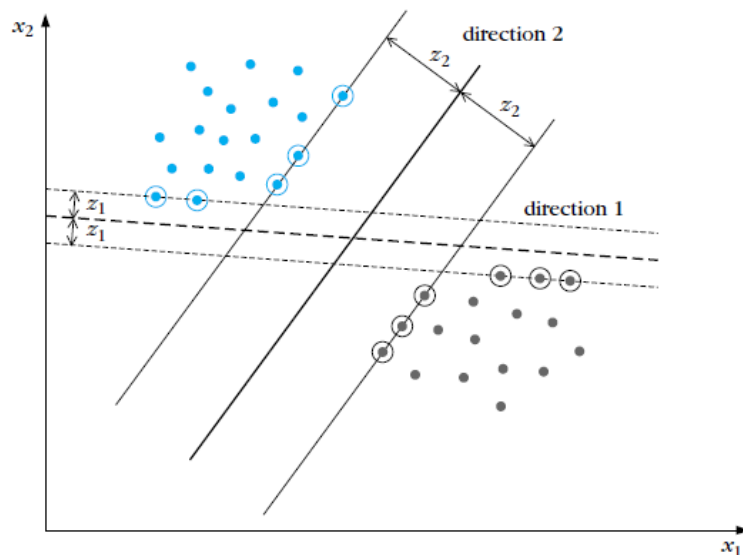
或者

$$\max_w d = \max(\|z_{\pm}\|) = \max_w \frac{1}{\|w\|}$$

线性支持向量机

目标函数

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2$$
$$s.t \quad y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0, \quad i=1,2,\dots,l$$



参考文献

1. Pattern Recognition 2nd. 《模式识别》(第二版), 边肇祺, 张学工等, 清华大学出版社, 2000.1。
2. Pattern Classification, 2nd. 模式分类, 第二版。
3. 机器学习, 周志华, 清华大学出版社, 2016.
4. 数据挖掘中的新方法-支持向量机, 邓乃扬, 田英杰著, 2004。