线性分类器

翟婷婷

扬州大学
信息工程（人工智能）学院
zhtt@yzu.edu.cn

2022年春

课程目标

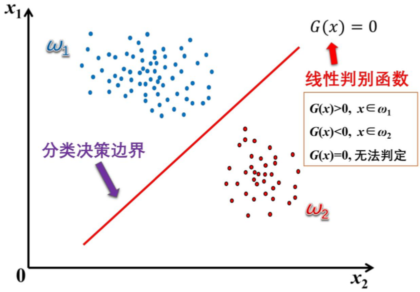
- 掌握线性判别函数的一般表达和参数的几何意义
- 理解线性可分和线性不可分问题
- 了解广义线性化的思想
- 掌握线性判别函数值的几何意义

线性判别函数

- 在d维的特征空间中，**线性判别函数**的一般形式为：

$$G(\boldsymbol{x}) = \mathbf{w}^{\mathsf{T}}\boldsymbol{x} + b$$

其中，$\mathbf{w} = [w_1, w_2, \cdots, w_d]^{\mathsf{T}}$称为权向量，$b$称为偏置。$(\mathbf{w}, b)$是判别函数的参数，需要通过自主学习得到。$G(\boldsymbol{x})$将特征空间划分为两个决策域：$G(\boldsymbol{x}) > 0$和$G(\boldsymbol{x}) < 0$，而$G(\boldsymbol{x}) = 0$是**决策边界方程**。



线性判别函数加上对应的分类决策规则，就构成**线性分类器**。

线性分类决策边界

- 在d维的特征空间中，线性分类决策边界的表达式为：

$$G(\boldsymbol{x}) = \mathbf{w}^{\intercal}\boldsymbol{x} + b = 0$$

- 当$d = 1$, 决策边界是一个点:

$$x = -\frac{b}{\mathrm{w}}$$

- 当$d = 2$, 决策边界是一条直线:

$$\mathrm{w}_1 x_1 + \mathrm{w}_2 x_2 + b = 0$$

- 当$d = 3$, 决策边界是一个平面:

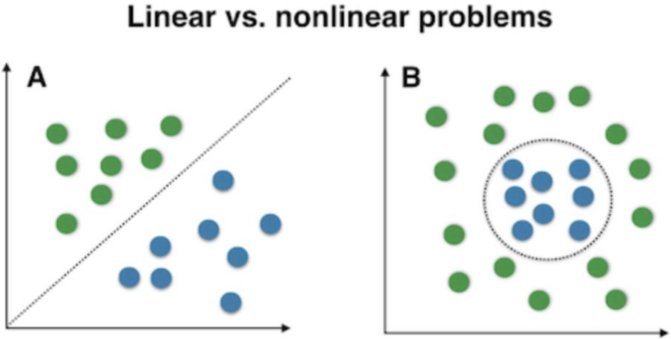
$$\mathrm{w}_1 x_1 + \mathrm{w}_2 x_2 + \mathrm{w}_3 x_3 + b = 0$$

- 当$d > 3$, 决策边界是一个超平面:

$$\mathbf{w}^{\intercal}\boldsymbol{x} + b = 0$$

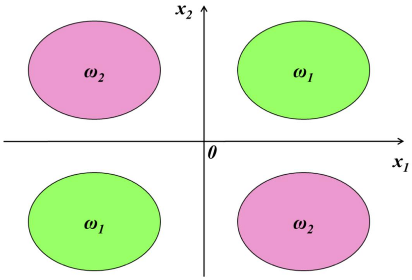
线性可分vs线性不可分问题

- 以二分类问题为例：如果一个二分类问题能够找到一个**线性判别函数$G(x)$**将两类实例分得开，那么该二分类问题是**线性可分的**，否则就是**线性不可分的**。



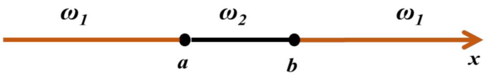
线性不可分问题

- 通过非线性判别函数进行分类:



$$\boldsymbol{x} = (x_1, x_2)$$

$$G(\boldsymbol{x}) = x_1 x_2$$
$$\boldsymbol{x} \in \omega_1, G(\boldsymbol{x}) > 0$$
$$\boldsymbol{x} \in \omega_2, G(\boldsymbol{x}) < 0$$

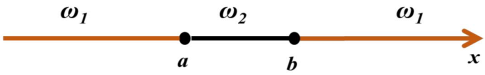


$$G(x) = (x - a)(x - b)$$
$$x \in \omega_1, G(x) > 0$$
$$x \in \omega_2, G(x) < 0$$

非线性判别函数的训练方法比线性判别函数的训练方法复杂的多!

广义线性化

- 线性可分问题往往比线性不可分问题更容易解决!
- 有没有方法能将**线性不可分**问题转化为**线性可分**问题呢?



$$G(x) = (x-a)(x-b)$$
$$G(x) > 0,\ x \in \omega_1$$
$$G(x) < 0,\ x \in \omega_2$$

$$G(x) = (x-a)(x-b) = x^2 - (a+b)x + ab$$

令$y_1 = x^2$, $y_2 = x$

$$G(x) = y_1 - (a+b)y_2 + ab\ = g(y_1, y_2)$$

二维特征空间中的线性判别函数！

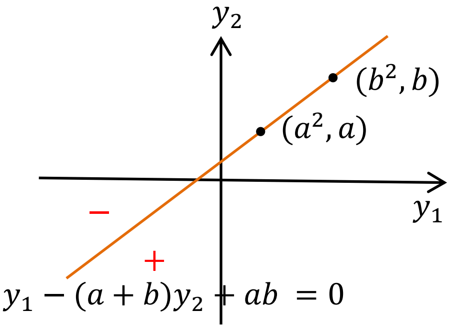
广义线性化

- 通过非线性变换，将实例从一维空间映射到二维空间中：

$$x \rightarrow \boldsymbol{y} = (y_1, y_2) = (x^2, x)$$

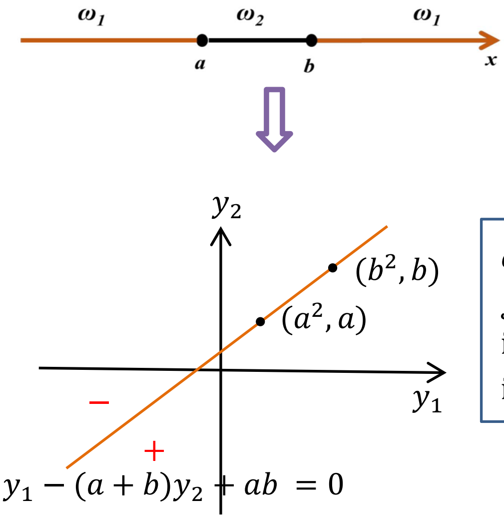
得到一个线性判别函数：

$$g(y_1, y_2) = y_1 - (a + b)y_2 + ab$$



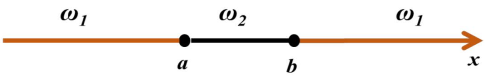
$$a \rightarrow (a^2, a),\quad g(a^2, a) = 0$$
$$b \rightarrow (b^2, b),\quad g(b^2, b) = 0$$

广义线性化

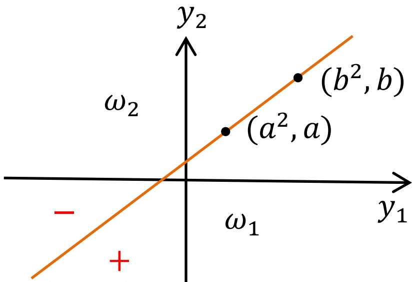


$$G(x) = (x-a)(x-b)$$
$$G(x) > 0, x \in \omega_1$$
$$G(x) < 0, x \in \omega_2$$

$$c \to (c^2, c)$$
$$g(c^2, c) = (c-a)(c-b)$$
$$\text{if } c \in \omega_1, \quad g(c^2, c) > 0$$
$$\text{if } c \in \omega_2, \quad g(c^2, c) < 0$$



$$G(x) = (x - a)(x - b)$$
$$G(x) > 0,\ x \in \omega_1$$
$$G(x) < 0,\ x \in \omega_2$$



$$y_1 - (a + b)y_2 + ab = 0$$

$$c \to (c^2, c)$$
$$g(c^2, c) = (c - a)(c - b)$$
$$\text{if } c \in \omega_1,\quad g(c^2, c) > 0$$
$$\text{if } c \in \omega_2,\quad g(c^2, c) < 0$$

广义线性化

- 通过一个非线性变换，我们将

  识别问题从　　**低维**特征空间 → **高维**特征空间

  线性不可分 → 线性可分

  这种方法就称为“广义线性化”。

- 是不是所有在低维空间中线性不可分的问题，映射到高维空间都可以变成线性可分的问题呢？需要映射到多少维才行呢？

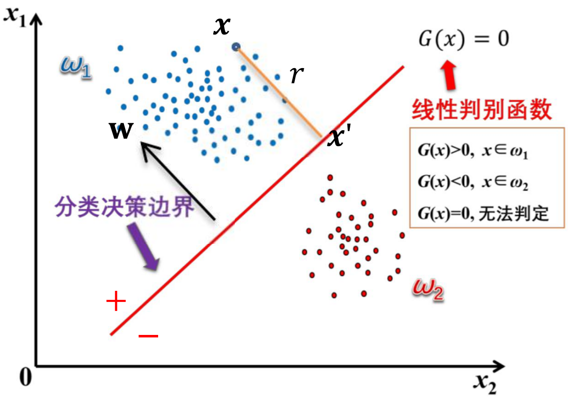
  我们后面讲到svm的时候再解答。

线性判别函数值的几何意义

- 在d维的特征空间中，线性判别函数的一般形式为：

$$G(\boldsymbol{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{x} + b$$

$\mathbf{w}$是超平面$\mathbf{w}^{\mathrm{T}}\boldsymbol{x} + b = 0$的法向量。 r为实例$\boldsymbol{x}$到决策边界的距离。



①$\boldsymbol{x}$在决策边界正侧

$$\boldsymbol{x}' + r \cdot \frac{\mathbf{w}}{||\mathbf{w}||} = \boldsymbol{x}$$

$$G(\boldsymbol{x}') = 0$$

$$\Rightarrow \ r = \frac{G(\boldsymbol{x})}{||\mathbf{w}||}$$

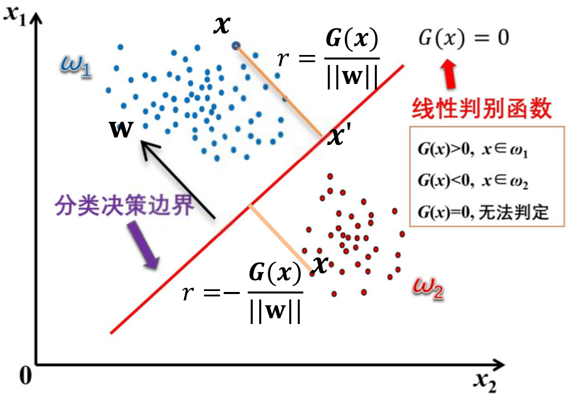
②$\boldsymbol{x}$在决策边界负侧

$$\boldsymbol{x}' - r \cdot \frac{\mathbf{w}}{||\mathbf{w}||} = \boldsymbol{x}$$

$$G(\boldsymbol{x}') = 0$$

$$\Rightarrow \ r = -\frac{G(\boldsymbol{x})}{||\mathbf{w}||}$$

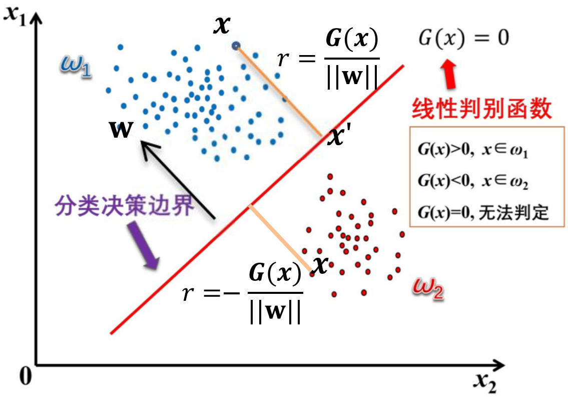
线性判别函数值的几何意义



$$r=\frac{|G(x)|}{||\mathbf{w}||}$$

- 判别函数$G(x)$的符号表示实例x位于决策边界的正侧还是负侧。实例x到决策边界的距离r正比于判别函数$G(x)$的绝对值。

线性判别函数值的几何意义



$$r = \frac{|G(x)|}{||\mathbf{w}||}$$

- **w**仅代表决策边界的法向量方向，其长度不会影响决策边界在特征空间中的位置，因此完全可以取$||\mathbf{w}|| = 1$，此时，**判别函数的绝对值就是实例到决策边界的距离。**

小结

- 线性可分的问题能够通过一个**线性判别函数**将两类的实例分得开，而线性不可分问题则不能。
- 在**低维**特征空间中**线性不可分**的问题，可以通过非线性变换，映射到**高维**特征空间中变成**线性可分**的问题，这称为“广义线性化”。

重点

- 线性判别函数的一般形式为：

$$G(\boldsymbol{x}) = \mathbf{w}^{\intercal}\boldsymbol{x} + b$$

其中$\mathbf{w}$称为权向量，b称为偏置。$\mathbf{w}^{\intercal}\boldsymbol{x} + b = 0$是一个超平面的方程，$\mathbf{w}$是该超平面的法向量。

小结

- 一个线性判别函数$G(x)$只能将特征空间划分为**两个**决策域：$G(x) > 0$ 和$G(x) < 0$，因此只能用于二分类问题。
- 线性判别函数值的几何意义是：

  实例$\boldsymbol{x}$的判别函数的绝对值，即$|G(\boldsymbol{x})|$，正比于该实例到分类决策边界$G(x) = 0$的距离。

重点