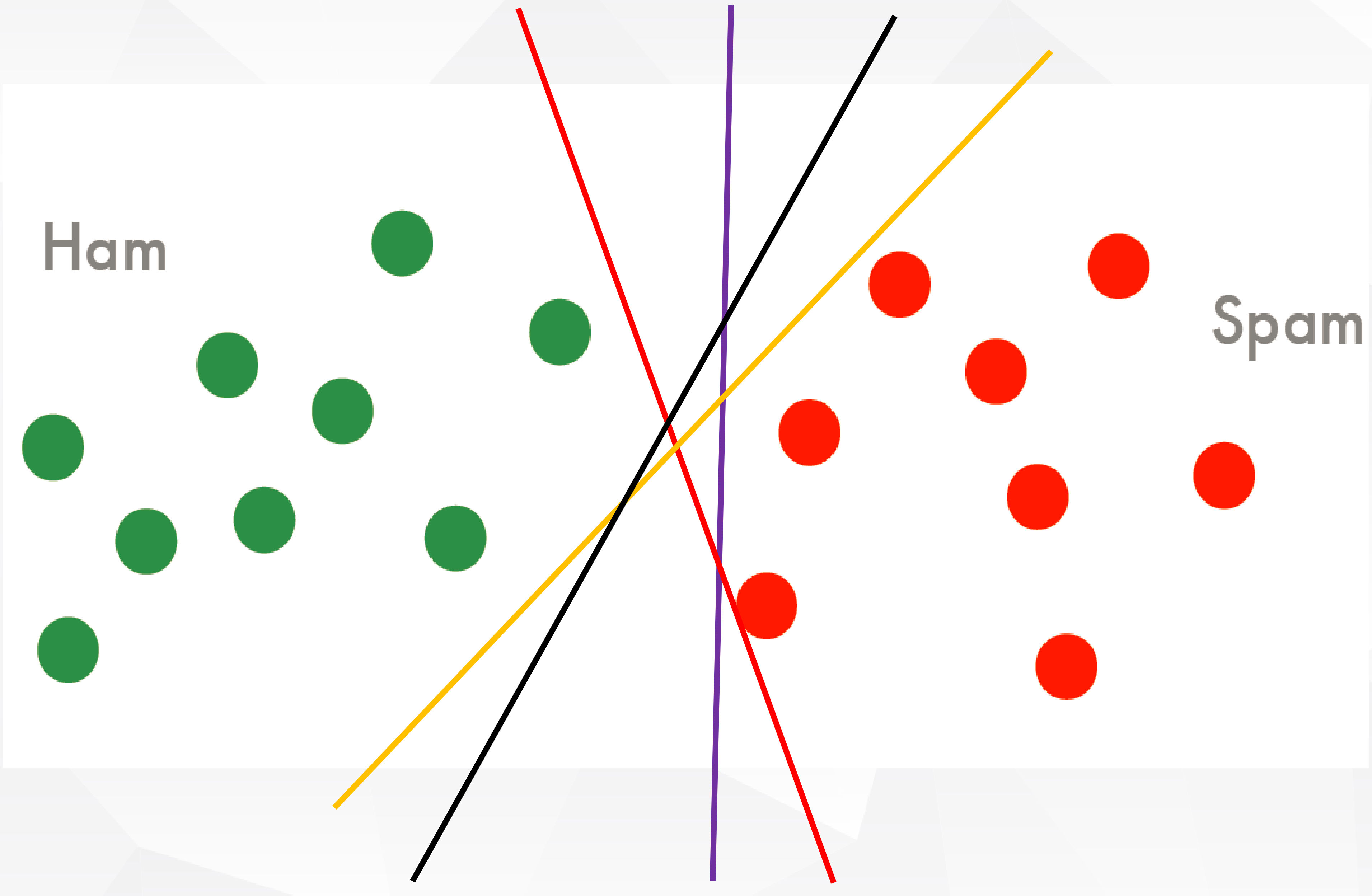
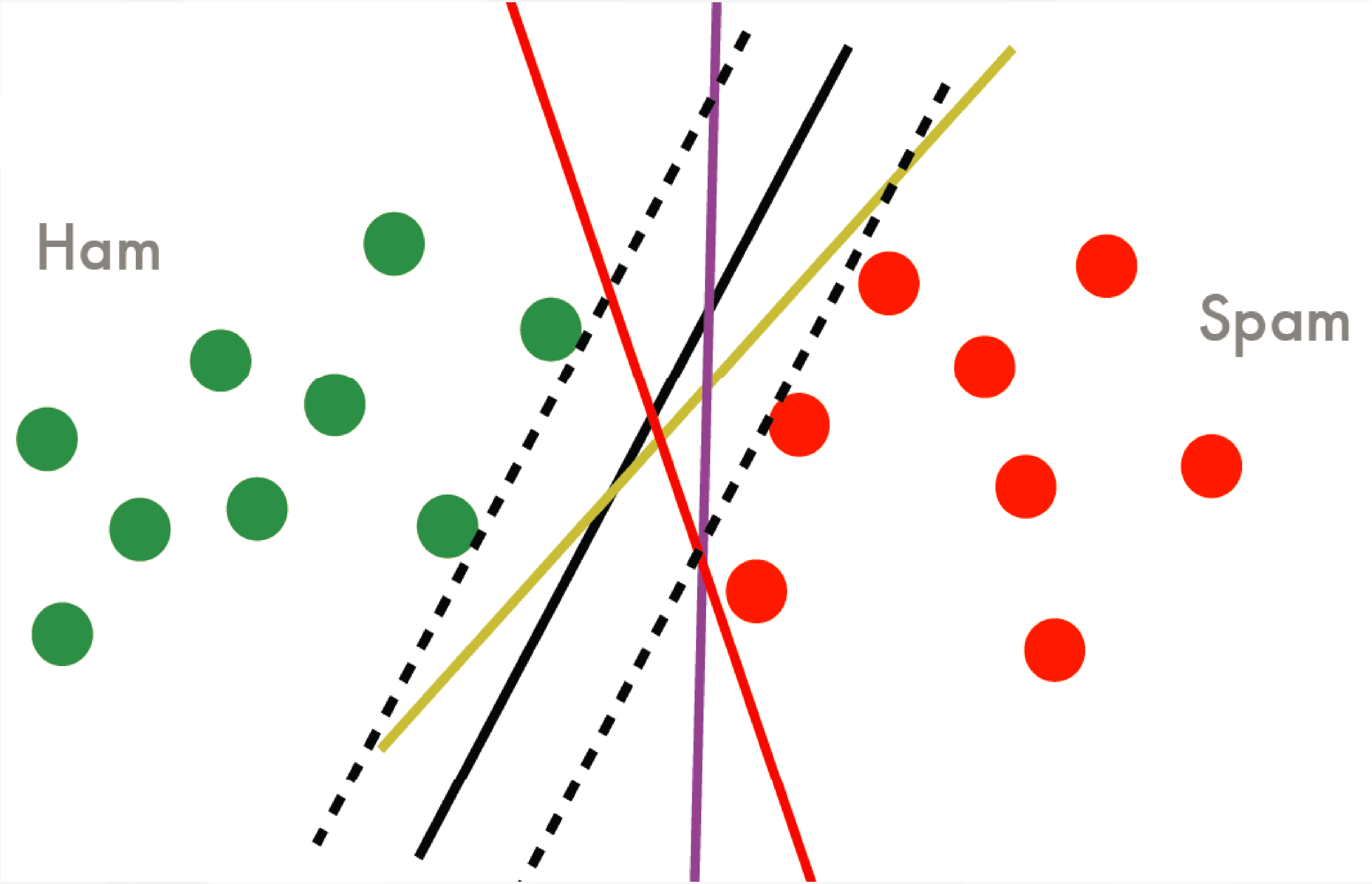
第七章：支持向量机 Support Vector Machine, SVM

- 线性支持向量机
- 核支持向量机
- 序列最小优化算法

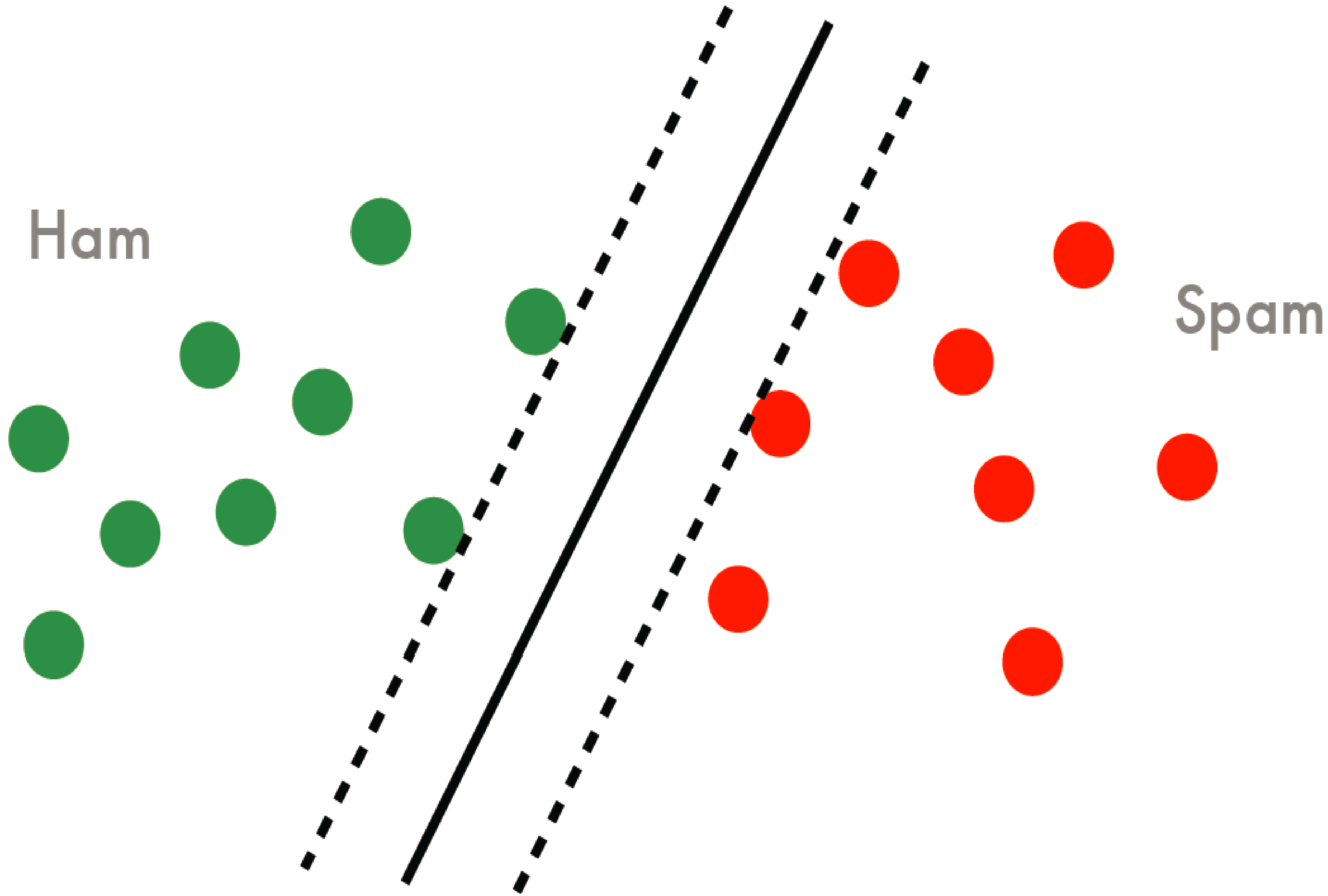
线性分类器



线性分类器



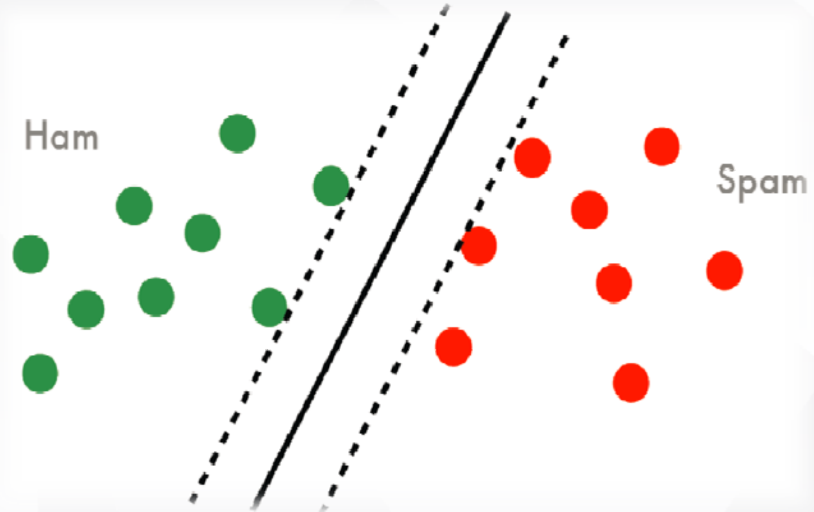
线性分类器



间隔 （Margin）

■Logistic 回归中

$$P(y=1|\boldsymbol{x})=\frac{1}{1+e^{-w^T\boldsymbol{x}}}$$

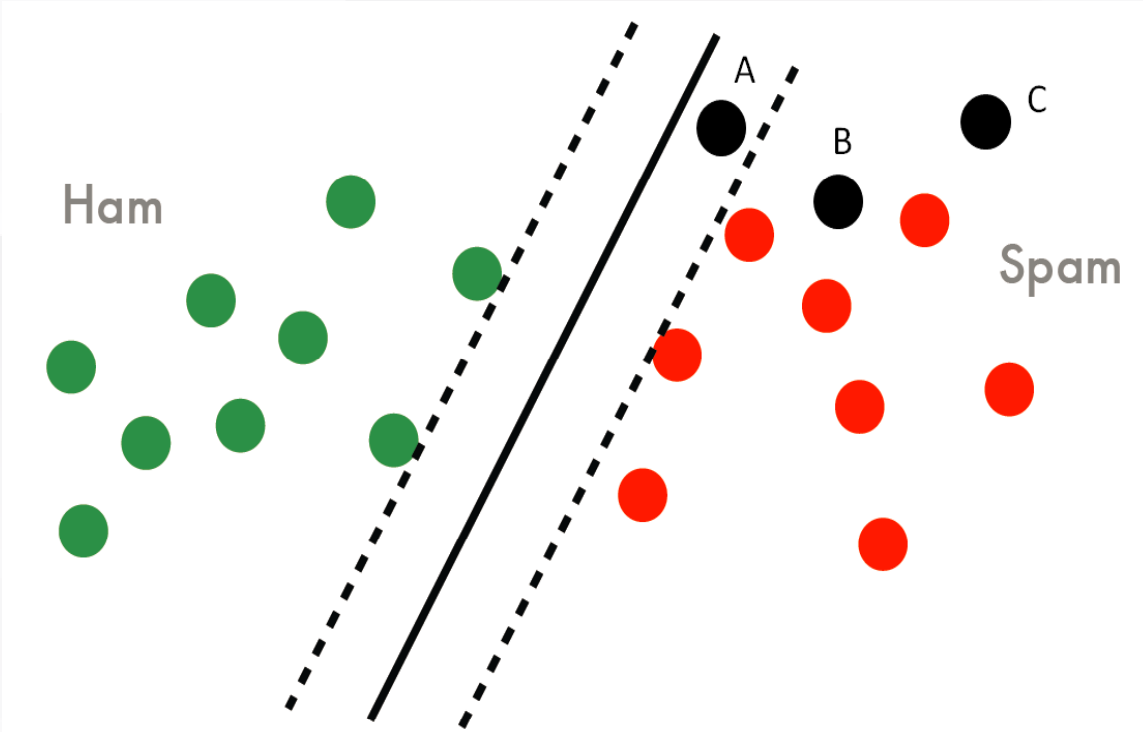


- $\boldsymbol{w}^T\boldsymbol{x}$ 越大，y=1的置信度越高

•给定训练数据，如果当y=1时， $\boldsymbol{w}^T\boldsymbol{x}\gg 0$ ；当y=-1时，$\boldsymbol{w}^T\boldsymbol{x}\ll 0$ ，则w是一个比较好的参数。

■问题: 除了概率，如何定义分类器的置信度?

间隔Margin



■将样本A、B、C分为哪一类？对于哪个样本的预测的置信度高?

点到分离超平面的距离反映了预测的置信度。

二分类问题-线性分类器

- **样本** $(\boldsymbol{x}, y), y \in \{-1, 1\}$
- 线性分类器:

$$f_{w,b}(\boldsymbol{x}) = g(\boldsymbol{w}^T \boldsymbol{x} + b) \qquad g(z) = \begin{cases} 1, & \text{如果 } z \geq 0 \\ -1, & \text{如果} z < 0 \end{cases}$$

- 超平面: $\boldsymbol{w}^T \boldsymbol{x} + b = 0$

函数间隔(Functional Margin)

- 对于一样训练样本 $(\boldsymbol{x}^i, y^i)$ ，它到 $(\boldsymbol{w}, b)$ 确定的超平面的函数间隔为：

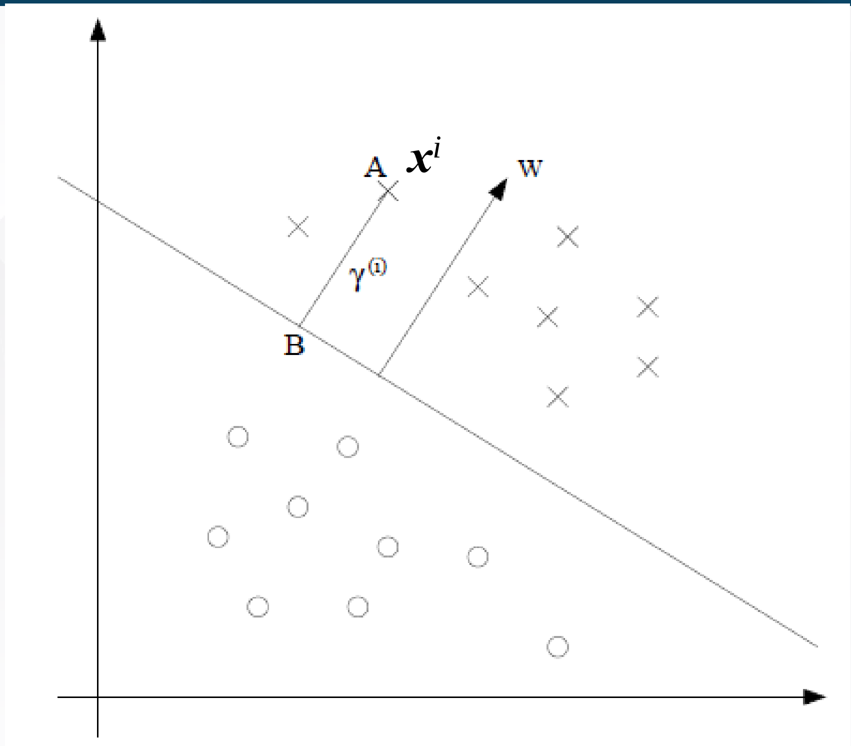
$$\hat{\gamma}^i = y^i(\boldsymbol{w}^T\boldsymbol{x}^i + b)$$

  - $y^i = 1$ ， $\boldsymbol{w}^T\boldsymbol{x}^i + b$ 是一个大的正数.
  - $y^i = -1$ ， $\boldsymbol{w}^T\boldsymbol{x}^i + b$ 是一个比较小的负数.
  - $y^i(\boldsymbol{w}^T\boldsymbol{x}^i + b) > 0$ ，模型对样本的预测是正确的.
  - 大的函数间隔→ 确信的、正确的预测.

- 对于训练数据集 $S = \{(\boldsymbol{x}^i, y^i), i = 1, \ldots, N\}$ ，它的函数间隔定义为

$$\hat{\gamma} = \min_i \hat{\gamma}^i, i = 1, 2, \ldots, N$$

几何间隔



■ 对于样本（x^i ,y^i） y^i= 1，它到决策界面的距离 γ^i 是线段AB的长度。

$$\gamma^i = ?$$

几何间隔

- 点B： $\boldsymbol{x}^i - \gamma^i \boldsymbol{w} / \|\boldsymbol{w}\|_2$
- 点B在决策边界上。

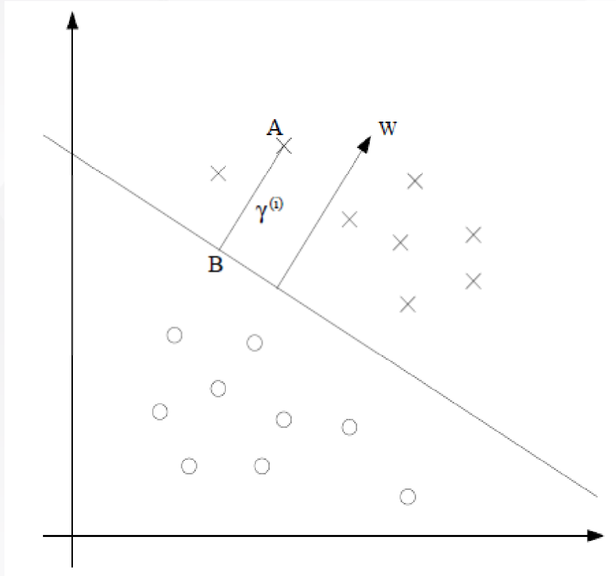
$$\boldsymbol{w}^T(\boldsymbol{x}^i - \gamma^i \boldsymbol{w}/\|\boldsymbol{w}\|_2) + b = 0$$

- 求解方程可得γ ：

$$\gamma^i = \frac{\boldsymbol{w}^T\boldsymbol{x}^i + b}{\|\boldsymbol{w}\|_2} = \left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}$$

- 样本 $(\boldsymbol{x}^i, y^i)$ 的几何间隔为：

$$\gamma^i = y^i\left(\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}\right)$$



几何间隔

- 训练数据集 $S=\{(\boldsymbol{x}^i,y^i),i=1,\ldots,N\}$ 关于判别界面(参数为 $(\boldsymbol{w},b)$ ）)的几何间隔：

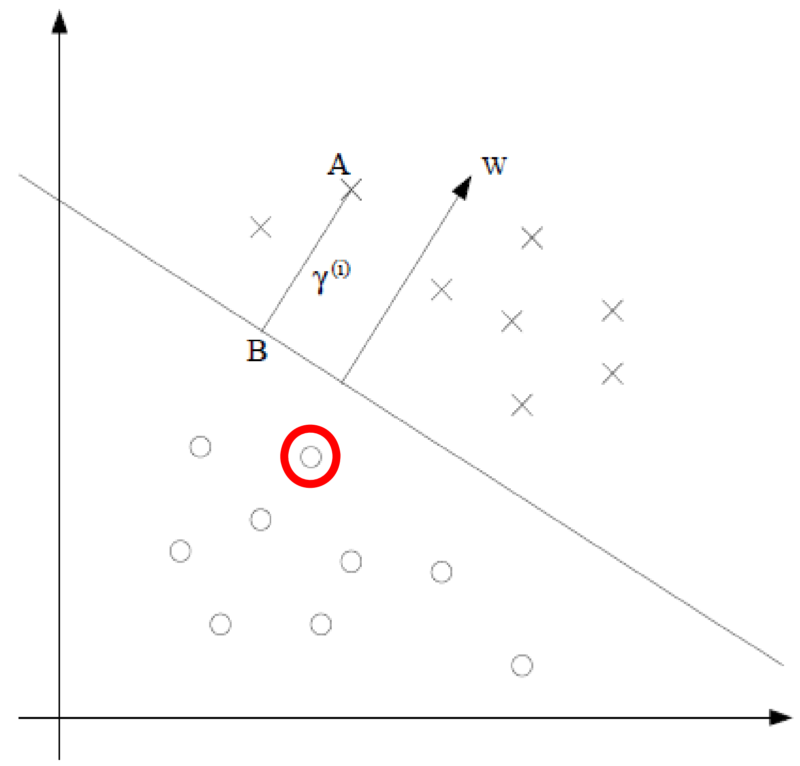
$$\gamma=\min_i \gamma^i, i=1,2,\ldots,N$$

- 函数间隔与几何间隔:

$$\gamma^i=\frac{\hat{\gamma}^i}{\|\boldsymbol{w}\|_2},\quad \gamma=\frac{\hat{\gamma}}{\|\boldsymbol{w}\|_2}.$$

•如果 $\|\boldsymbol{w}\|_2=1$ ， 二者相等。

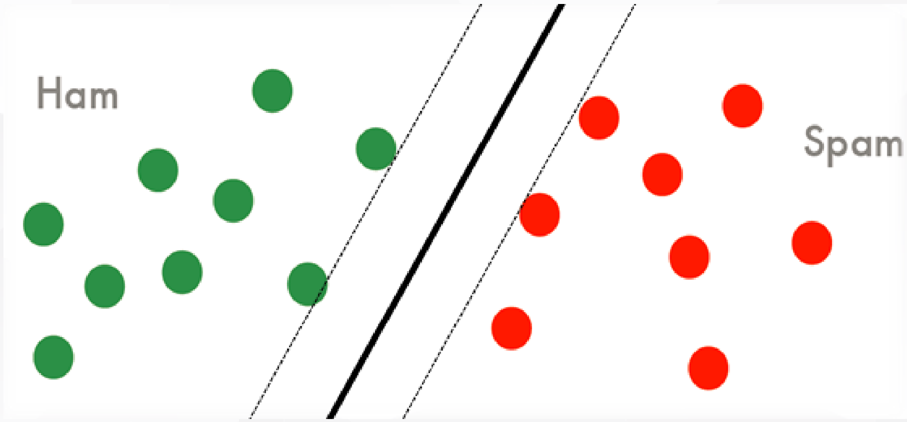
- **几何间隔具有不变性.**



最优间隔分类器

- 给定一个训练集，一个自然的想法是试图找到一个使**几何间隔最大化**的决策边界，这表示对训练集的有可信的预测并且对训练数据的良好“拟合”。

- 假设训练数据是线性可分的。



- 间隔最大化 ：

$$\max_{\gamma,\boldsymbol{w},b} \ \gamma$$

$$s.t. \ \ y^i\left(\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}\right) \geq \gamma, i = 1,\ldots,N$$

最优间隔分类器

- 问题转换:

$$\max_{\gamma,w,b} \frac{\hat{\gamma}}{\|\boldsymbol{w}\|_2}$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right) \geq \hat{\gamma}, i = 1,\dots,N$$

- 简化问题 $\hat{\gamma}=1$

$$\min_{w,b} \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right) \geq 1, i = 1,\dots,N$$

- 可以使用现成的**二次规划程式**求解（QP）。
  - •例 如Python中Cvxopt.solvers.qp(P,q,G,h,A,b)， matlab中的函数是quadprog

二次规划(QP, Quadratic Programming)定义：目标函数为二次函数，约束条件为线性约束，属于最简单的一种非线性规划。

线性SVM

- 输入：线性可分的训练数据集 $S=\{(\boldsymbol{x}^i, y^i), i=1,\ldots,N\}$
- 输出：判别函数及决策/判别界面。
  - 通过求解如下最优化问题来得到最优分类器的参数 $(\boldsymbol{w}^*, b)$
  -

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\geq 1, i=1,\ldots,N$$

  - 分离超平面: $(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*=0$
  - 判别函数： $f_{\boldsymbol{w},b}(\boldsymbol{x})=sign((\boldsymbol{w}^*)^T\boldsymbol{x}+b^*)$

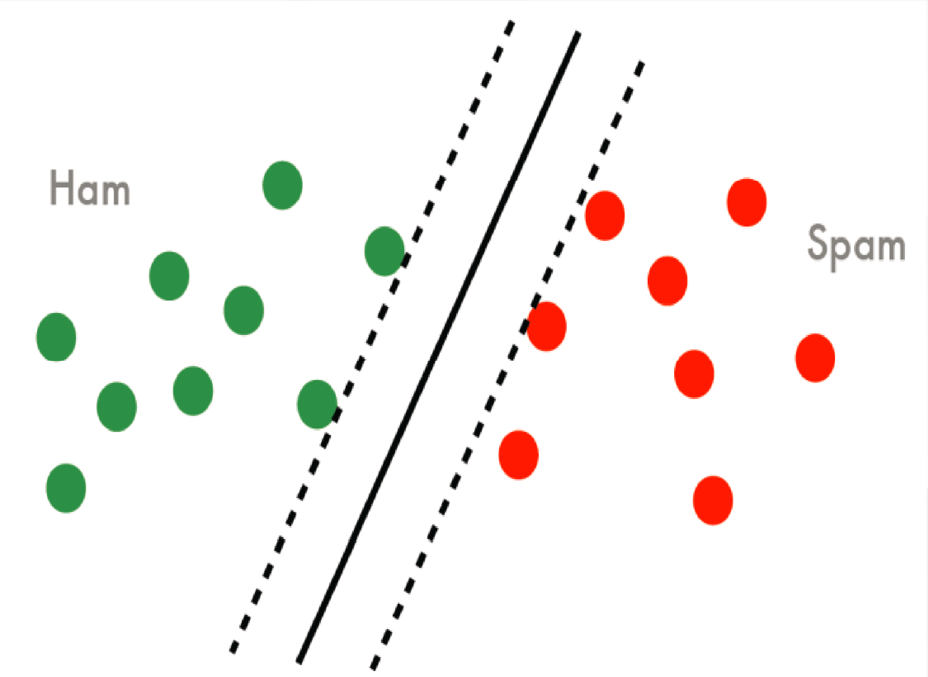
理论保证：对于线性可分的训练数据集，最大间隔分类器存在且唯一。

支持向量和间隔

- **支持向量：** 距分离超平面最近的训练样本。

$$y\left(\boldsymbol{w}^T\boldsymbol{x}+b\right)=1$$

- 函数间隔: $\hat{\gamma}=1$
- 几何间隔: $\gamma=\frac{1}{\|\boldsymbol{w}\|_2}$
- 间隔: $\frac{2}{\|\boldsymbol{w}\|_2}$



线性SVM -作业

■给定如下训练数据集。

$\boldsymbol{x}^1=[3\ 3]^T, \boldsymbol{x}^2=[4\ 3]^T, y^1=1, y^2=1$

$\boldsymbol{x}^3=[1\ 1], y^3=-1$

通过求解SVM的原始问题来求解最大间隔的分离超平面。

拉格朗日对偶：简单情形

■等式约束的最优化问题:

$$\min_{w} f(\boldsymbol{w})$$
$$s.t. \quad h_i(\boldsymbol{w}) = 0, \ i = 1, \dots, l$$

•拉格朗日函数：

$$L(\boldsymbol{w}, \boldsymbol{\beta}) = f(\boldsymbol{w}) + \sum_{i=1}^{l} \boxed{\beta_i} h_i(\boldsymbol{w})$$

拉格朗日乘子

■求解：偏导数为0:

$$\frac{\partial L}{\partial w_i} = 0, \quad \frac{\partial L}{\partial \beta_i} = 0$$

拉格朗日对偶: 一般情形

■约束条件中含有等式及不等式约束时:

$$\begin{aligned}&\min_{w}\ f(w)\\&s.t.\ \ g_i(w)\le 0,\ i=1,\dots,k\\&\qquad h_i(w)=0,\ i=1,\dots,l\end{aligned}$$

原问题
Primary Problem

■广义拉格朗日函数:

非负

$$L(w,\alpha,\beta)=f(w)+\sum_{i=1}^{k}\alpha_i g_i(w)+\sum_{i=1}^{l}\beta_i h_i(w)$$

拉格朗日乘子

拉格朗日对偶：一般情形

■定义如下最优化问题:

$$\theta_P(\boldsymbol{w}) = \max_{\alpha,\beta:\alpha_i\geq 0} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

■如 $\boldsymbol{w}$ 违反了约束条件:

$$\boxed{\begin{aligned} g_i(\boldsymbol{w}) &\leq 0,\ i=1,\dots,k \\ h_i(\boldsymbol{w}) &= 0,\ i=1,\dots,l \end{aligned}}$$

$$\theta_P(\boldsymbol{w}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i>0} f(\boldsymbol{w}) + \sum_{i=1}^{k}\alpha_i g_i(\boldsymbol{w}) + \sum_{i=1}^{l}\beta_i h_i(\boldsymbol{w}) =$$

■如果满足所有约束条件:

$$\theta_P(\boldsymbol{w}) = ?$$

拉格朗日对偶: 一般情形

■分析下面的最优化问题:

$$\min_{w} \theta_P(\boldsymbol{w}) = \min_{w} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

■原始问题:

$$\boxed{\begin{aligned} &\min_{w} \ f(\boldsymbol{w}) \\ &s.t. \ \ g_i(\boldsymbol{w}) \leq 0, \ \ i = 1,\ldots,k \\ &\qquad\ h_i(\boldsymbol{w}) = 0, \ \ i = 1,\ldots,l \end{aligned}}$$

■定义该问题的最优值：

$$p^* = \min_{w} \theta_P(\boldsymbol{w})$$

原问题的最优值

拉格朗日对偶: 一般情形

■对偶问题：$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_{\alpha,\beta:\alpha_i\geq 0} \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$

$$\theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

■对偶问题的最优值：

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

■原问题与对偶问题：

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \min_{\boldsymbol{w}} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta}) = p^*$$

拉格朗日对偶: 什么时候$d^*=p^*$?

■ 如果原问题是凸优化问题，且存在严格满足约束条件的w。

$$\min_{w} f(w)$$
$$s.t.\ \ g_i(w) \le 0,\ i=1,\dots,k$$
$$h_i(w)=0,\ i=1,\dots,l$$

$f(.)$、$g_i(.)$是凸函数，$h_i(.)$是仿射函数

$g_i(w) < 0$

■ 则存在 w^*, α^*, β^* ，使得 w^* 是原问题的解，α^*, β^* 是对偶问题的解，且满足:

$$p^* = d^* = L(w^*, \alpha^*, \beta^*)$$

■ w^*, α^*, β^* 满足Karush-Kuhn-Tucker (KKT) 条件:

$$\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0,\ i=1,2,\dots,M$$
$$g_i(w^*) \le 0,\ i=1,\dots,k$$
$$h_i(w^*) = 0,\ i=1,\dots,l$$
$$\boxed{\alpha_i^* g_i(w^*) = 0,\ i=1,\dots,k}$$
$$\alpha_i^* \ge 0,\ i=1,\dots,k$$

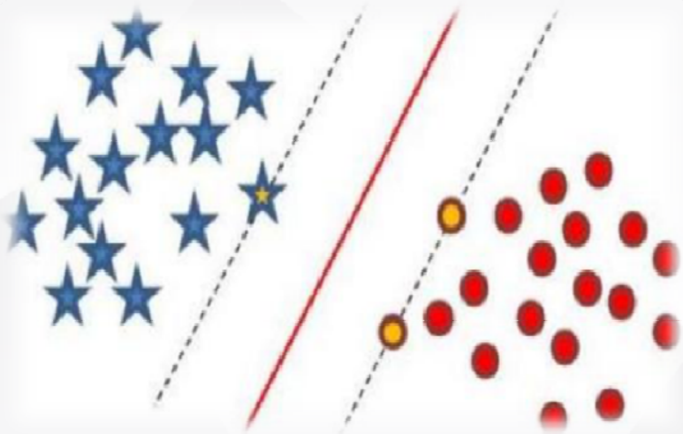
KKT 对偶互补性

最优间隔分类器：对偶解

■ 最优间隔分类器——原问题：

$$\min_{w,b} \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right) \geq 1, i = 1, \ldots, N$$



■ 约束条件：

$$g_i(\boldsymbol{w}) = -y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right) + 1 \leq 0,\ i = 1, \ldots, N$$

■ KKT对偶互补条件：$\alpha_i^* g_i(\boldsymbol{w}^*) = 0,\ i = 1, \ldots, k$

$$g_i(\boldsymbol{w}^*) < 0 \Rightarrow \alpha_i^* = 0$$

$$\boxed{\alpha_i^* \neq 0 \Rightarrow g_i(\boldsymbol{w}^*) = 0}$$

支持向量

■ 支持向量的数量远小于训练样本的数目。

最优间隔分类器: 对偶解

■拉格朗日函数:

$$L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|_2^2-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1\right]$$

■对偶解:

•固定 $\boldsymbol{\alpha}$， 关于参数$\boldsymbol{w}$ 和 b 最小化 $L(\boldsymbol{w},b,\boldsymbol{\alpha})$ 得到 $\theta_D(\boldsymbol{\alpha})$

$$\theta_D(\boldsymbol{\alpha})=\min_{\boldsymbol{w},b} L(\boldsymbol{w},b,\boldsymbol{\alpha})$$

•最大化 $\theta_D(\boldsymbol{\alpha})$ 得到对偶问题的最优值 d^*

最优间隔分类器: 对偶解

■ 固定 $\boldsymbol{\alpha}$， 关于参数 $\boldsymbol{w}$ 和 b 最小化 $L(\boldsymbol{w},b,\boldsymbol{\alpha})$ 得到 $\theta_D(\boldsymbol{\alpha})$

- 求解 $\boldsymbol{w}$ 和 b

$$\boxed{L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|_2^2-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1\right]}$$

$$\nabla_w L(\boldsymbol{w},b,\boldsymbol{\alpha})=\boldsymbol{w}-\sum_{i=1}^{N}\alpha_i y^i \boldsymbol{x}^i=0,$$

$$\Downarrow$$

$$\boldsymbol{w}=\sum_{i=1}^{N}\alpha_i y^i \boldsymbol{x}^i$$

$$\frac{\partial}{\partial b}L(\boldsymbol{w},b,\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i y^i=0,$$

- 代入拉格朗函数

$$\theta_D(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^i y^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$$

最优间隔分类器: 对偶解

■对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y^i y^j \alpha_i \alpha_j \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j$$

$$s.t.\ \alpha_i \geq 0, i = 1, \ldots, N,$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0.$$

■假设 $\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_l^*)$ 是对偶问题的最优解，原问题的解是?

■分类超平面是？判别函数是?

线性可分SVM (对偶)

- 输入：线性可分的训练数据集 $S=\{(\boldsymbol{x}^i, y^i), i=1,\ldots,N\}$
- 输出：分离超平面和判别函数.
  - 通过求解对偶问题来得到最优解 $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_l^*)$

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y^i y^j \alpha_i \alpha_j \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j$$

$$s.t.\ \alpha_i \geq 0, i=1,\ldots,N,$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0.$$

  - 得到原问题的最优解 $(\boldsymbol{w}^*, b^*)$

$$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y^i \boldsymbol{x}^i,\ b^* = y^j - \sum_{i=1}^{N} \alpha_i^* y^i \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j.\quad \alpha_j^* > 0$$

- 分离超平面：$(\boldsymbol{w}^*)^T \boldsymbol{x} + b^* = 0$ ，判别函数：$f_{w,b}(\boldsymbol{x}) = (\boldsymbol{w}^*)^T \boldsymbol{x} + b^*.$

支持向量

$$w^* = \sum_{i=1}^{N} \alpha_i^* y^i \boldsymbol{x}^i,\ b^* = y^j - \sum_{i=1}^{N} \alpha_i^* y^i \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j.\quad \alpha_j > 0$$

- 计算w^* 时 b^* ，只需要利用 $\alpha_i > 0$ 的那些样本(支持向量)来计算。

$$f_{w,b}(\boldsymbol{x}) = \left(\boldsymbol{w}^*\right)^T \boldsymbol{x} + b^* = \sum_{i=1}^{N} \alpha_i^* y^i \left(\boldsymbol{x}^i\right)^T \boldsymbol{x} + b^*$$

- 预测时是否需要所有的训练样本来计算?

- 对偶形式：只需计算训练样本与输入特征的内积.

$$\left(\boldsymbol{x}^i\right)^T \boldsymbol{x} = <\boldsymbol{x}^i, \boldsymbol{x}>$$

- **核技巧!**

线性可分SVM-作业

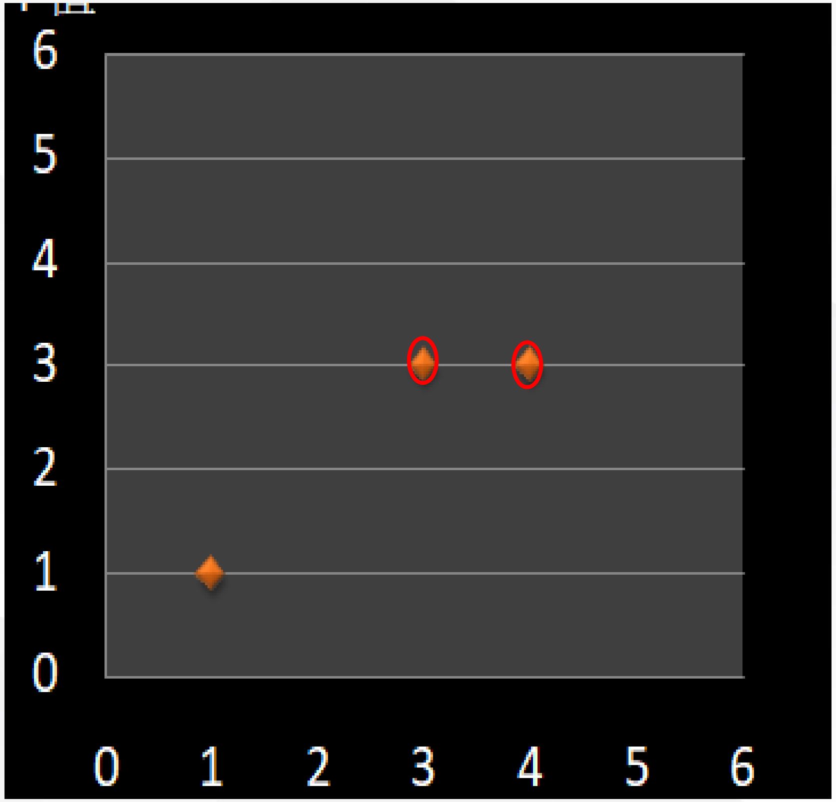
- 训练数据

$x^1=[3\ 3]^T, x^2=[4\ 3]^T,$

$y^1=1, y^2=1$

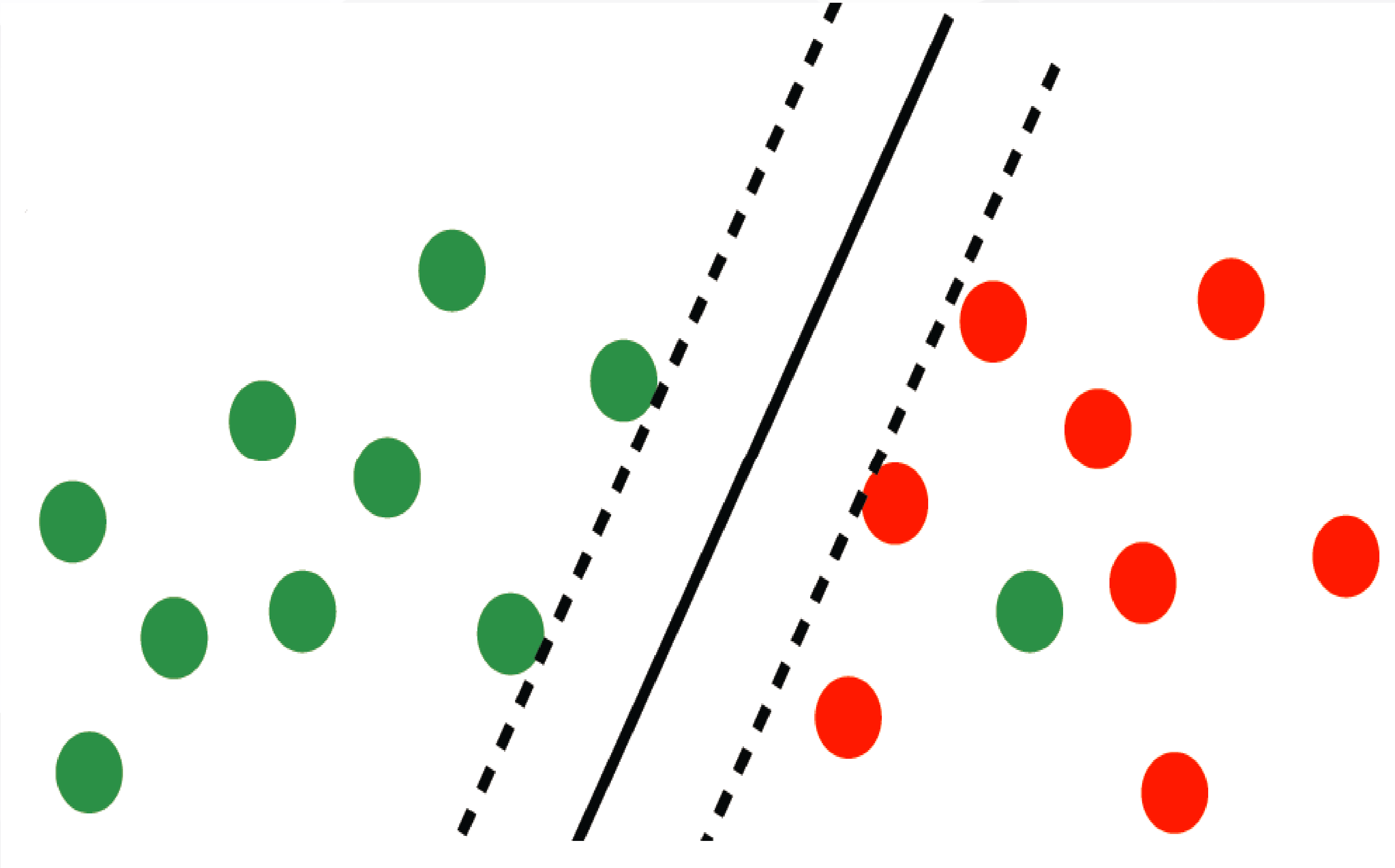
$x^3=[1\ 1], y^3=-1$

通过求解对偶问题来得到最大间隔分类器，写出分类超平面及判别函数。



软间隔（Soft Margin）分类器

■如果训练数据是线性不可分的，怎么办?

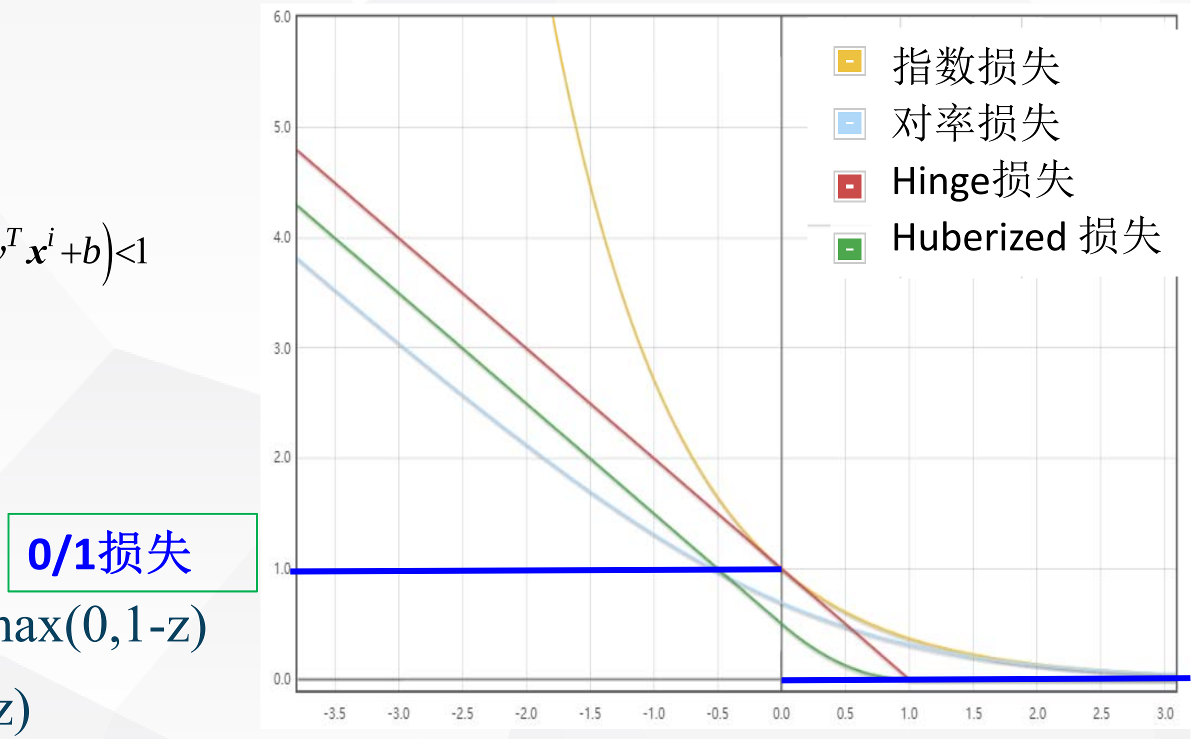


软间隔（Soft Margin）分类器

- 允许一些样本（离群点或噪声样本）违反原来的不等式约束条件：$y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\geq 1$，但是数目要少。
- 目标函数：

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}I_{y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)<1}$$

- 不易求解!
- 替代损失：



hinge损失：$l_{hinge}(z)=\max(0,1-z)$

指数损失: $l_{exp}(z)=\exp(-z)$

Huberized损失： $l_{hub}(z)=0.5-z,\ z<0$；$0.5(1-z)^2, 0\leq z\leq 1$；$0,\ z>1$

对率损失(logistic): $l_{log}(z)=\log(1+\exp(-z))$

软间隔（Soft Margin）分类器

■Hinge 损失:

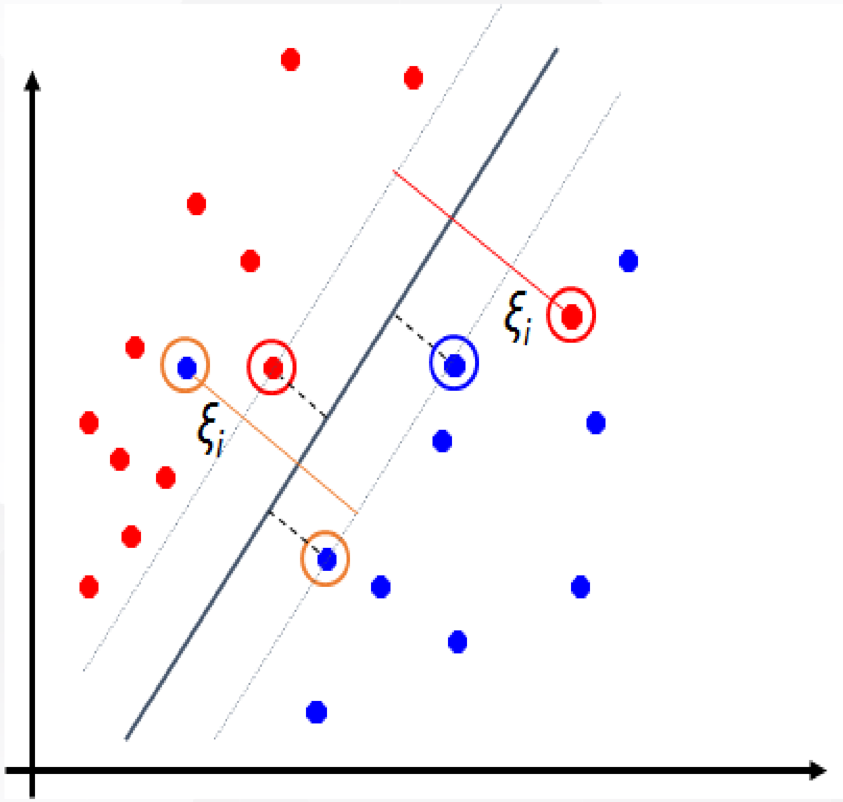
$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N} \max\left\{0, 1 - y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right)\right\}$$

■松弛变量 ξ_i :

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N} \xi_i$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i + b\right) \geq 1 - \xi_i,\ i = 1, \ldots, N,$$

$$\xi_i \geq 0,\ \ i = 1, \ldots, N.$$



■软间隔分类器.

■凸二次规划问题。

软间隔（Soft Margin）分类器—对偶问题

■拉格朗日函数 $L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\boldsymbol{\eta})=\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}\xi_i$

$$-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1+\xi_i\right]-\sum_{i=1}^{N}\eta_i\xi_i.$$

■固定 $\boldsymbol{\alpha}$、$\boldsymbol{\eta}$，关于参数$\boldsymbol{w}$、b 和 ξ，最小化 $L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\boldsymbol{\eta})$，得到 $\theta_D(\boldsymbol{\alpha},\boldsymbol{\eta})$

$$\theta_D(\boldsymbol{\alpha},\eta)=\min_{\boldsymbol{w},b,\xi}L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\eta)\quad
\begin{cases}
\nabla_{\boldsymbol{w}}L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\boldsymbol{\eta})=\boldsymbol{w}-\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i=0, \Rightarrow \boxed{\boldsymbol{w}=\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i}\\
\frac{\partial}{\partial b}L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\boldsymbol{\eta})=\sum_{i=1}^{N}\alpha_i y^i=0,\\
\frac{\partial}{\partial \xi_i}L(\boldsymbol{w},b,\xi,\boldsymbol{\alpha},\boldsymbol{\eta})=C-\alpha_i-\eta_i=0.
\end{cases}$$

$$\theta_D(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^i y^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$$

■最大化 $\theta_D(\boldsymbol{\alpha},\boldsymbol{\eta})$，得到最优值 d^*。

软间隔（Soft Margin）分类器—对偶问题

■**对偶问题**

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y^i y^j \alpha_i \alpha_j \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j$$

$$s.t.\ 0 \le \alpha_i \le C, i = 1,..,N,$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0.$$

■假设 $\boldsymbol{\alpha}^* = (\alpha_1^*, ..., \alpha_l^*)$ 是对偶问题的最优解，原问题的解是?

■分类超平面是？分类决策函数是?

非线性可分 SVM (对偶问题)

- 输入： 训练数据集 $S=\{(\boldsymbol{x}^i,y^i),i=1,\dots,N\}$
- 输出：分类超平面与分类决策函数.
  - 选择参数C， 并求解对偶问题，得最优解 $\boldsymbol{\alpha}^*=(\alpha_1^*,\dots,\alpha_l^*)$

$$\max_{\boldsymbol{\alpha}}\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j$$

$$s.t.\ 0\le\alpha_i\le C,i=1,..,N,$$

$$\sum_{i=1}^{N}\alpha_iy^i=0.$$

  - 求解最优的 $(\boldsymbol{w}^*,b^*)$ $:\boldsymbol{w}^*=\sum_{i=1}^{N}\alpha_i^*y^i\boldsymbol{x}^i,\ b^*=y^j-\sum_{i=1}^{N}\alpha_i^*y^i\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$ $\boxed{0<\alpha_j^*<C}$

  - 分类超平面: $\left(\boldsymbol{w}^*\right)^T\boldsymbol{x}+b^*=0$ ，判别函数: $f_{w,b}(\boldsymbol{x})=\left(\boldsymbol{w}^*\right)^T\boldsymbol{x}+b^*.$

非线性SVM：支持向量

- 支持向量：$\alpha_i^* > 0$.

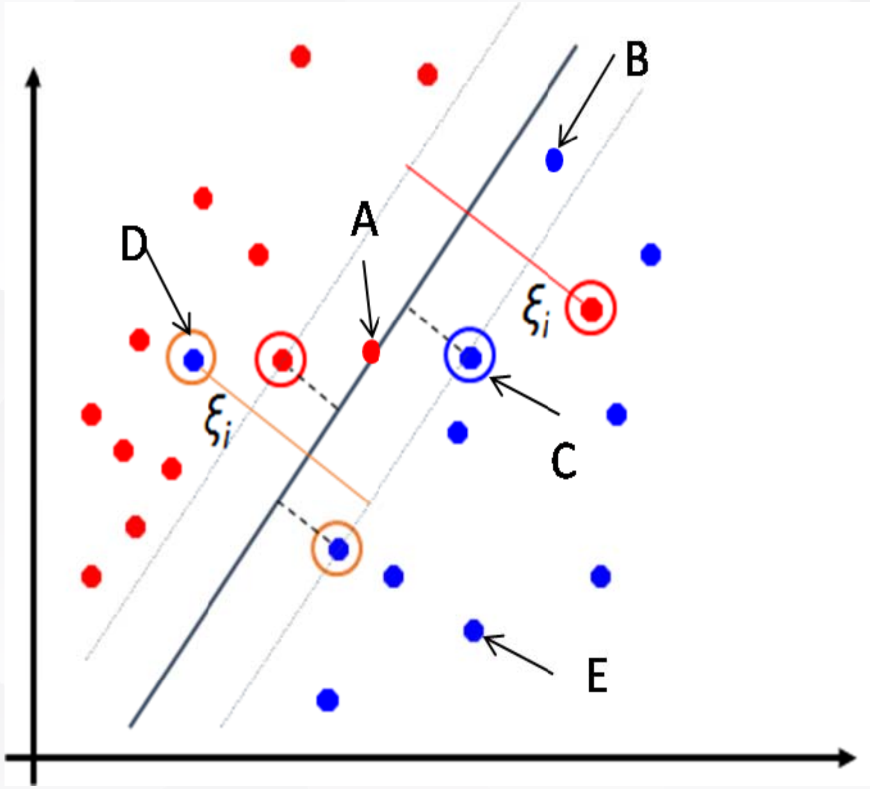
- KKT对偶互补条件：$\boxed{\alpha_i^* g_i(\boldsymbol{w}^*) = 0}$ $\boxed{\eta_i^* \xi_i = 0.}$ $\boxed{\frac{\partial}{\partial \xi_i} L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\eta}) = C - \alpha_i - \eta_i = 0.}$

$$\alpha_i^* \left[y^i \left((\boldsymbol{w}^*)^T \boldsymbol{x}^i + b \right) - 1 + \xi_i \right] = 0, \ \eta_i^* \xi_i = 0.$$

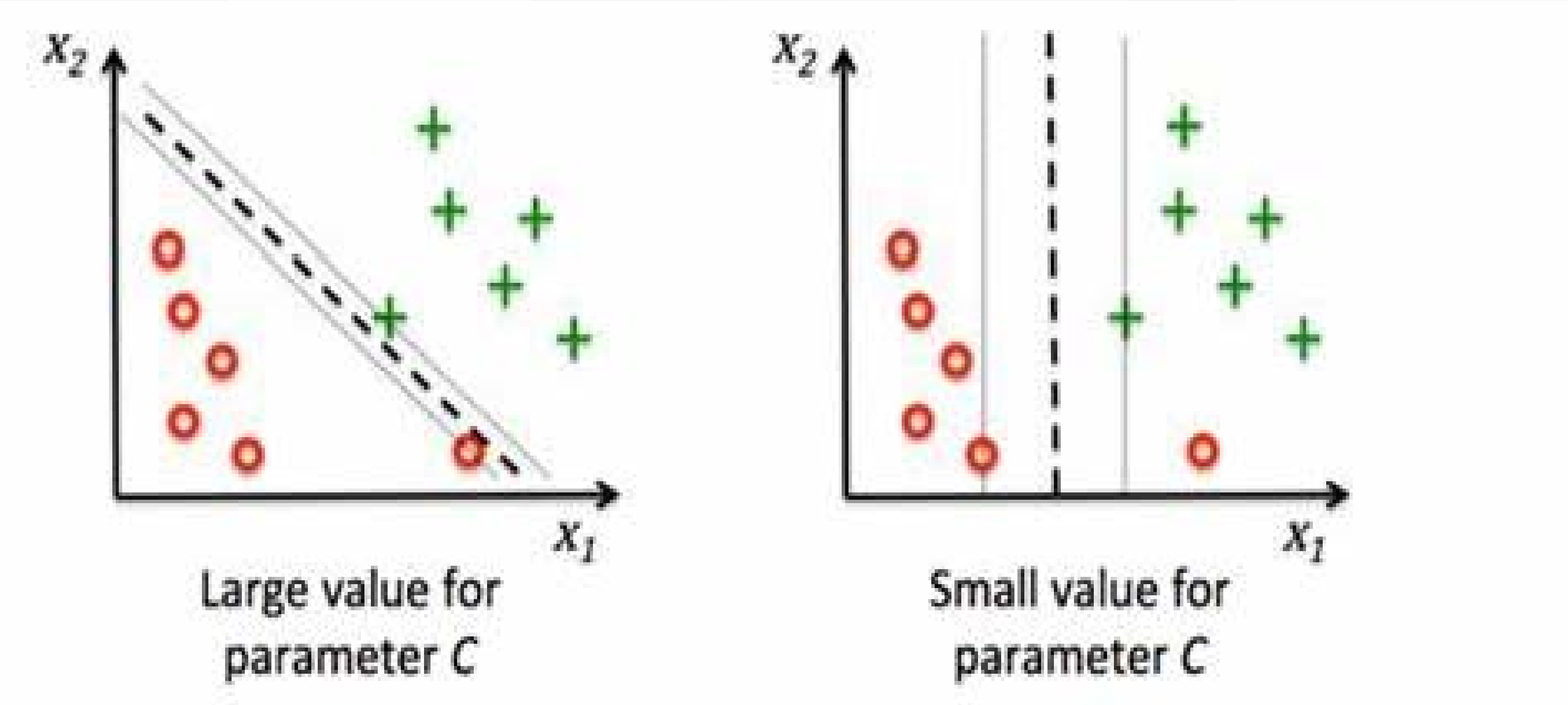
- 如果 $\alpha_i^* = 0$ ，$y^i \left((\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) \geq 1$

- 如果 $0 < \alpha_i^* < C$ ，$y^i \left((\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) = 1$

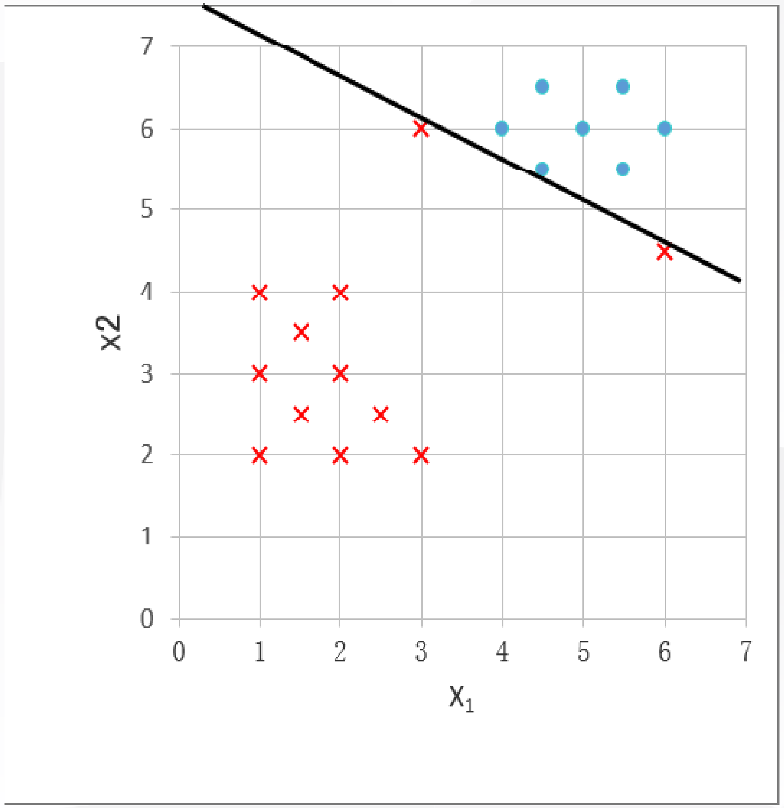
- 如果 $\alpha_i^* = C$ ，$y^i \left((\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) \leq 1$
  - $0 < \xi_i < 1$
  - $\xi_i = 1$
  - $\xi_i > 1$

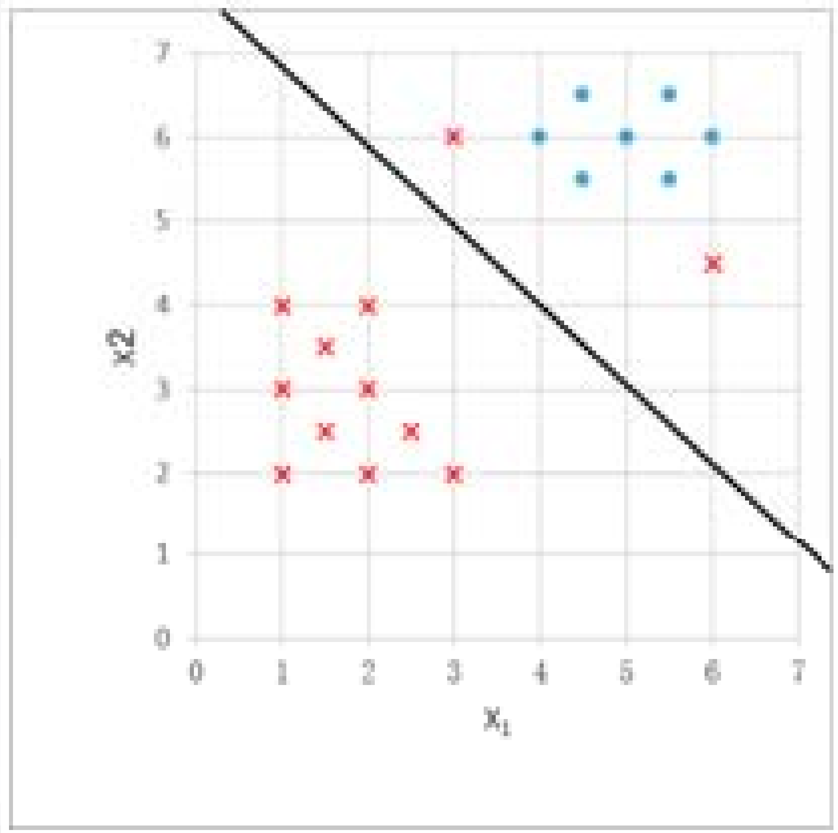


参数 C


x_2
x_1
Large value for
parameter C
x_2
x_1
Small value for
parameter C

参数 C





手写数字识别实验 – KL+linear-SVM

在Minist数据集上进行PCA分析

图像数据维数高，而且特征之间（像素之间）相关性很高，因此我们预计用很少的维数就能保留足够多的信息

```
#导入必要的工具包
import pandas as pd
import numpy as np
from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
import time
from scipy.io import loadmat
```

```
train = pd.read_csv('./data/MNIST_train.csv').values
test = pd.read_csv('./data/MNIST_test.csv').values
#train
```

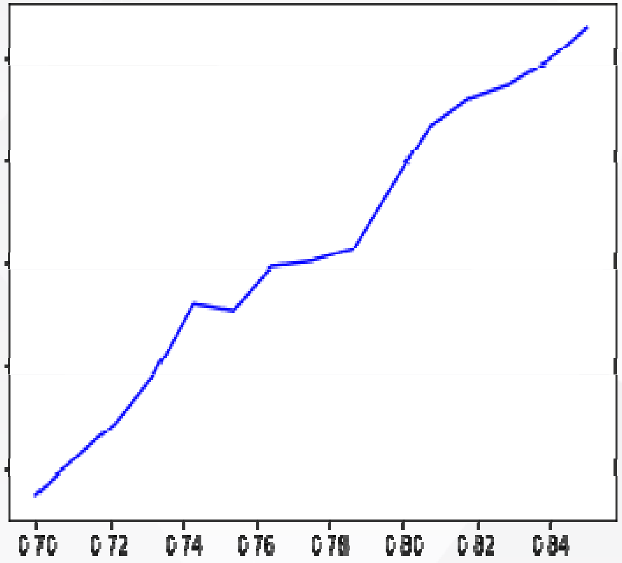
```
    pca = PCA(n_components=n)
    print("PCA begin with n_components: {}".format(n));
    pca.fit(X_train)

    # 在训练集和测试集降维
    X_train_pca = pca.transform(X_train)
    X_val_pca = pca.transform(X_val)

    # 利用SVC训练
    print('SVC begin')
    #kernel : 核函数，默认是rbf，可以是'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'
    clf1 = svm.SVC(kernel='linear')
    clf1.fit(X_train_pca, y_train)

    # 返回accuracy
    accuracy = clf1.score(X_val_pca, y_val)

    end = time.time()
    print("accuracy: {}, time elaps:{}".format(accuracy, int(end-start)))
    return accuracy
```



保留的信息量
降维后，特征维度越高，保留的信息量越多

- 10类 准确率：0.98209820

作业

■推导软间隔SVM的对偶形式。