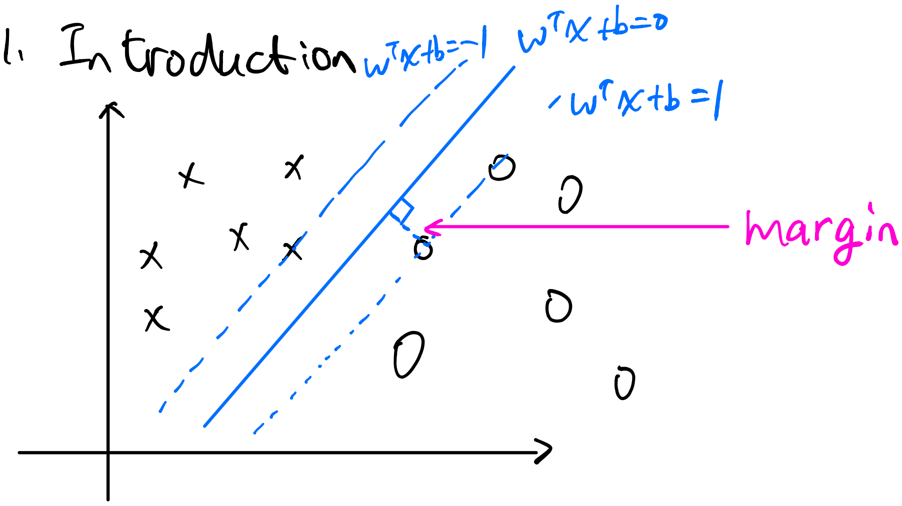
SVM

1. Introduction



① **Hyperplane:**

$$\begin{cases} w^Tx+b>0 \Rightarrow y=+1 \\ w^Tx+b<0 \Rightarrow y=-1 \end{cases} \Longrightarrow y_i(w^Tx_i+b)>0,\ \forall i=1,\cdots,N$$

② Find maximum margin

margin: 离超平面最近的点的距离

- Margin

点到超平面距离公式

$$\text{margin}(w,b)=\min_{\substack{w,b,\ x_i \\ i=1,\cdots,N}} \frac{1}{\|w\|}|w^Tx_i+b|$$

- Maximum margin

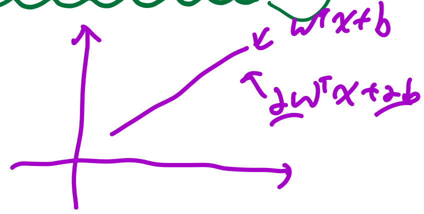
Explanation: 通过约束 w, b, 找到距离最近的点 x_i 以及距离 distance, 再变动 w, b, 找到 distance 最大的那个 w 和 b, 即找到最~~好的~~ hyperplane.

$$\Rightarrow \max_{w,b} \min_{\substack{x_i \\ i=1,\cdots,N}} \frac{1}{\|w\|}|w^Tx_i+b|$$

$$\Rightarrow \begin{cases} \max\limits_{w,b} \dfrac{1}{\|w\|} \min\limits_{\substack{x_i \\ i=1,\cdots,N}} |w^Tx_i+b| \Rightarrow \max\limits_{w,b} \dfrac{1}{\|w\|} \boxed{\min\limits_{\substack{x_i \\ i=1,\cdots,N}} y_i(w^Tx_i+b)} = 1 \\ \text{s.t. } y_i(w^Tx_i+b)>0 \Rightarrow \exists\, r>0, \text{使} \min y_i(w^Tx_i+b)=r \end{cases}$$

$r=1$

确定下来



$$\Rightarrow \begin{cases} \min\limits_{w,b} \|w\| \\ \text{s.t. } \min y_i(w^Tx_i+b)=1 \end{cases}$$

convex optimization.

$$\Rightarrow \begin{cases} \min\limits_{w,b} \underline{\frac{1}{2} w^T w} \text{ 只是方便计算} \\ s.t \quad y_i(w^T x_i + b) \geq 1, \ i=1,\cdots N \end{cases}$$

等价

③ Solve

$$\cdot \begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w \\ s.t. \quad 1 - y_i(w^T x + b) \leq 0 \end{cases}$$

带约束

定义 Lagrange function:

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \lambda_i (1 - y_i(w^T x + b))$$

上述目标则可转化为

prime problem

$$\begin{cases} \min\limits_{w,b} \max\limits_{\lambda} \mathcal{L}(w,b,\lambda) \\ s.t. \quad \lambda_i \geq 0 \end{cases}$$

无约束

Explanation:

① 弱对偶问题 $\min\limits_{\lambda} \max\limits_{w,b} \ell(w,b,\lambda) \geq \max\limits_{w,b} \min\limits_{\lambda} \ell(w,b,\lambda)$

② 强对偶问题 $\min\limits_{\lambda} \max\limits_{w,b} \ell(w,b,\lambda) = \max\limits_{w,b} \min\limits_{\lambda} \ell(w,b,\lambda)$

③ 什么时候满足强对偶条件?

满足 KKT条件 (后面讨论)

dual problem

$$\Rightarrow \begin{cases} \max\limits_{\lambda} \min\limits_{w,b} \mathcal{L}(w,b,\lambda) \\ s.t. \quad \lambda \geq 0 \end{cases}$$

开始求解上述 dual problem. 可分成 3 步:

1) $\boxed{\text{固定}\lambda}$, 要让 $\mathcal{L}$ 关于 w,b 最小化, 则分别对 w,b 求偏导:

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial(\overset{①}{\frac{1}{2} w^T w} + \sum_{i=1}^{N} \lambda_i (1 - y_i(w^T x_i + b)))}{\partial w} = 0$$

$$\Rightarrow w - \frac{\partial(\sum_{i=1}^{N} \lambda_i y_i x_i w^T)}{\partial w} ② = 0$$

$$\Rightarrow w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0$$

矩阵求导公式:

① $\frac{d x^T x}{dx} = 2x$

② $\frac{d x^T}{dx} = I$

N

$$\Rightarrow \boxed{w = \sum_{i=1} \lambda_i y_i x_i} \quad ①$$

$$\frac{\partial L}{\partial b} = \frac{\partial\left(\frac{1}{2}w^T w + \sum_{i=1}^{N} \lambda_i (1 - y_i(w^T x_i + b))\right)}{\partial b} = 0$$

$$\Rightarrow \frac{\partial\left(-\sum_{i=1}^{N} \lambda_i y_i b\right)}{\partial b} = 0$$

$$\Rightarrow \boxed{\sum_{i=1}^{N} \lambda_i y_i = 0} \quad ②$$

将①.②代入 $\mathcal{L}$，其中 $\mathcal{L} = \underbrace{\frac{1}{2}w^T w}_{2} + \underline{\sum_{i=1}^{N} \lambda_i (1 - y_i(w^T x_i + b))}$ 1

$$= \frac{1}{2}w^T w + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N} \lambda_i y_i w^T \overset{①}{x_i} - \sum_{i=1}^{N} \overset{②}{\lambda_i y_i} b$$

$$= \frac{1}{2}w^T w - w^T w + \sum_{i=1}^{N} \lambda_i$$

$$= -\frac{1}{2}w^T w + \sum_{i=1}^{N} \lambda_i$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i x_i\right)^T \left(\sum_{i=1}^{N} \lambda_i y_i x_i\right) + \sum_{i=1}^{N} \lambda_i$$

Explanation：y_i 是数，λ_i 被固定了，当成是数，所以 y_i, λ_i 不参与转置.

$$= -\frac{1}{2} \sum_{i=1}^{N} \lambda_i y_i (x_i)^T \sum_{j=1}^{N} \lambda_j y_j x_j + \sum_{i=1}^{N} \lambda_i$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i y_i (x_i)^T \lambda_j y_j x_j + \sum_{i=1}^{N} \lambda_i$$

则问题化为：

$$\max_{\lambda} \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$$\text{s.t.} \quad \lambda_i \geq 0$$

$$\sum_{i=1}^{N} \lambda_i y_i = 0, \quad i = 1, \cdots, N$$

原、对偶问题具有强对偶关系 $\Longleftrightarrow$ 满足KKT条件

本题中，KKT条件为

$$\begin{cases} \frac{\partial L}{\partial w} = 0, \quad \frac{\partial L}{\partial b} = 0, \quad \frac{\partial L}{\partial \lambda} = 0 \\ \lambda_i (1 - y_i(w^T x_i + b)) = 0 \quad a \end{cases}$$

$$\begin{cases} \lambda_i \geq 0 \\ 1-y_i(w^T x_i+b) \leq 0 \end{cases}$$

从而，可解得 $\boxed{w^* = \sum_{i=0}^{N} \lambda_i y_i x_i}$　　b

假设 $\exists (x_k, y_k)$. s.t. a.

则 $1-y_k(w^T x_k+b)=0$

$\Rightarrow \quad y_k(w^T x_k+b)=1$

$\Rightarrow \quad y_k^2(w^T x_k+b)=y_k$　　$y_k=\pm 1,\ y_k^2=1$

$\Rightarrow \quad w^T x_k+b=y_k$

$\Rightarrow \quad \boxed{b^* = y_k - \sum_{i=0}^{N} \lambda_i y_i x_i x_k}$

即最后的超平面为 $w^{*T} x + b^*$

后面再推导

2) 最后，剩拉格朗日乘子入需要求解，需要用到SMO算法。解决的问题就是在参数$[\lambda_1, \lambda_2, \cdots, \lambda_n]^T$上求最大值的问题，其中 x_i, y_i 均是已知数

$$\boxed{\begin{aligned} \max \quad & \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \\ \text{s.t.} \quad & a_i \geq 0,\ i=1,\cdots,n \\ & \sum_{i=1}^{N} \lambda_i y_i = 0 \end{aligned}}$$

补充上述推导过程中的数学知识

- 约束优化问题

参考资料

视频：凌肖 31-38

书：Boyd 第五章

Primal Problem

$$\begin{cases} \min\limits_{x \in R} f(x) \\ \text{s.t. } m_i(x) \leq 0 \quad i=1,\cdots,m \\ \quad\ \ n_j(x) = 0 \quad j=1,\cdots,n \end{cases}$$

why?

证：如果 x 违反了约束 $m_i(x)$，$m_i(x)>0$，则 $\max\limits_{\lambda} \mathcal{L} \to +\infty$

如果 x 符合 $m_i(x)\le 0$，则 $\max\limits_{\lambda} \mathcal{L} \ne +\infty$

$$\min_{x} \max_{\lambda} \mathcal{L} = \min_{x} \{\underset{(\text{符合的}x)}{\max \mathcal{L}}, \underset{(\text{不符合的}x)}{+\infty}\} = \min_{x} \max_{\lambda} \mathcal{L}$$

Lagrange Function.

Lagrange multiplier

$$\mathcal{L}(x, \lambda, \eta) = f(x) + \sum_{i=1}^{m} \lambda_i m_i + \sum_{j=1}^{n} \eta_j n_j$$

$$p \leftarrow \begin{cases} \min\limits_{x} \max\limits_{\lambda,\eta} \mathcal{L}(x,\lambda,\eta) \\ s.t.\quad \lambda_i \geq 0 \end{cases}$$

$\Rightarrow x \in \{$符合的 x 的集合$\}$

（问题是关于 x 的函数）

Dual Problem

$$d \leftarrow \begin{cases} \max\limits_{\lambda,\eta} \min\limits_{x} \mathcal{L}(x,\lambda,\eta) \\ s.t.\quad \lambda_i \geq 0 \end{cases}$$

（问题是关于 λ, η 的函数）

- 弱对偶性：对偶问题 $\le$ 原问题，即 $d \le p$

即证：$\max\limits_{\lambda,\eta} \min\limits_{x} \mathcal{L}(x,\lambda,\eta) \le \min\limits_{\lambda,\eta} \max\limits_{x} \mathcal{L}(x,\lambda,\eta)$

证：
$$\underbrace{\min_{x} \mathcal{L}(x,\lambda,\eta)}_{A(\lambda,\eta)} \le \mathcal{L}(x,\lambda,\eta) \le \underbrace{\max_{\lambda,\eta} \mathcal{L}(x,\lambda,\eta)}_{B(x)}$$

即 $A(\lambda,\eta) \le B(x)$

$\Rightarrow \max A(\lambda,\eta) \le \min B(x)$

$\Rightarrow \boxed{\max \min \mathcal{L}(x,\lambda,\eta) \le \min \max\limits_{\lambda,\eta} \mathcal{L}(x,\lambda,\eta)}$

• 对偶关系的几何解释

$$\begin{cases} \min\limits_{x \in R^n} f(x) \\ s.t. \quad m_1(x) \le 0 \end{cases}, \quad D: 定义域 \quad D = \operatorname{dom} f \cap \operatorname{dom} m_1$$

$\Rightarrow \mathcal{L}(x, \lambda) = f(x) + \lambda m_1(x), \quad \lambda \ge 0$

则 $p^* = \min f(x)$ （原问题的最优解）

$d^* = \max\limits_{\lambda} \min\limits_{x} \mathcal{L}(x, \lambda)$ （对偶最优解）

定义集合 $G = \{(\underset{\downarrow}{m_1(x)}, \underset{\downarrow}{f(x)}) \mid x \in D\}$ G: 二维坐标的一个区域

$= \{(u, t) \mid x \in D\}$

用G将 p^*, d^* 表示:

$p^* = \inf \{ t \mid (u, t) \in G, u \le 0 \}$ inf: 下确界

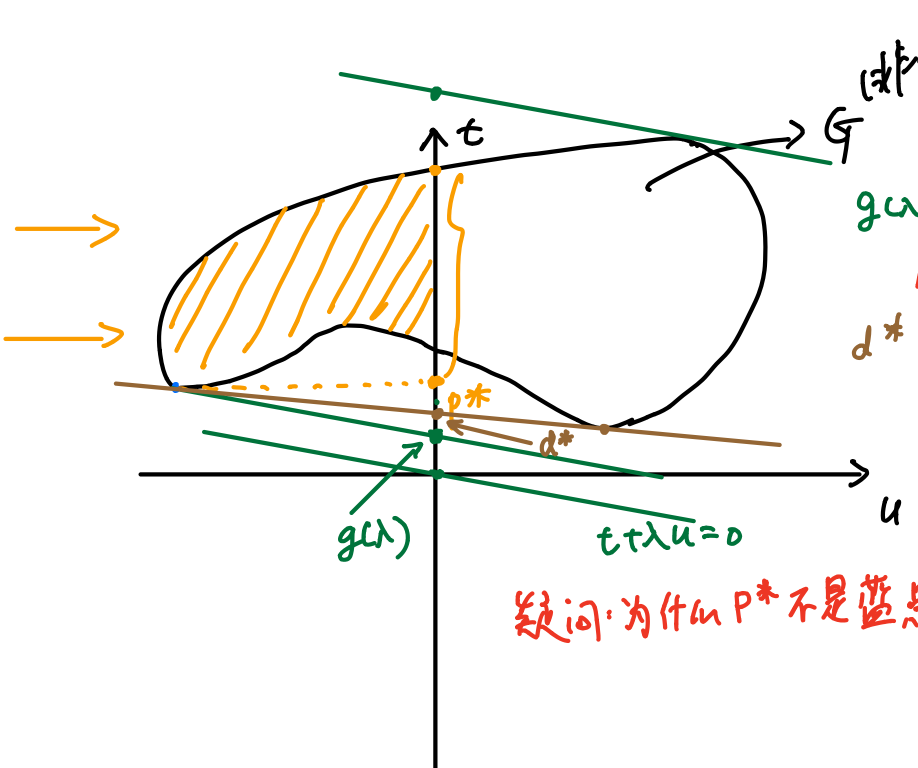
$d^* = \max\limits_{\lambda} \min\limits_{x} \underline{\mathcal{L}(x, \lambda)} \rightarrow f(x) + \lambda m_1(x)$

$= \max\limits_{\lambda} \underbrace{\min\limits_{x} (t + \lambda u)}_{g(\lambda)}$

$= \max\limits_{\lambda} g(\lambda)$

$g(\lambda) = \inf \{ t + \lambda u \mid (u, t) \in G \}$

问题：p^*、$g(\lambda)$、d^* 如何在图上表示？



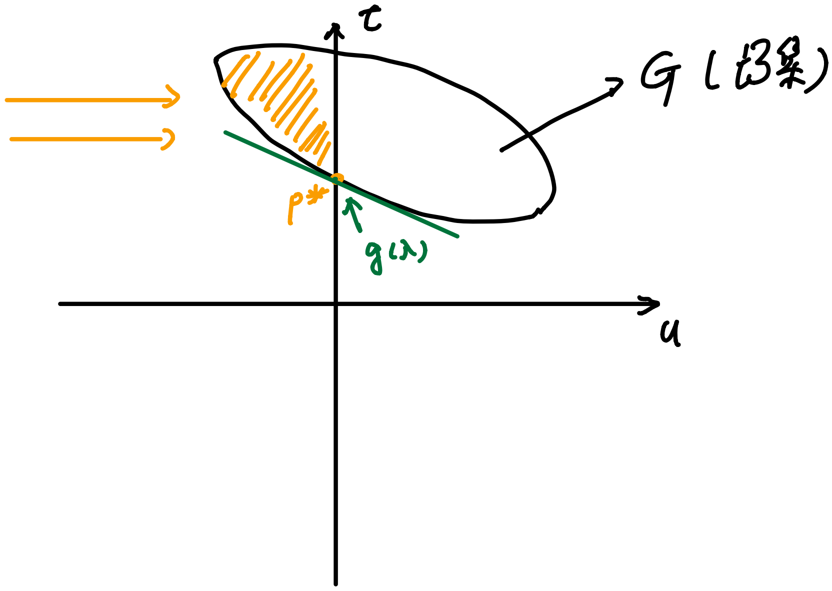
$p^* = \inf\{t \mid (u, t) \in G, u \le 0\}$

$g(\lambda) = \underline{\inf} \{ \underline{t + \lambda u} \mid (u, t) \in G \}$
必须相切 直线

$d^* = \max\limits_{\lambda} g(\lambda)$ → 两点相切时，$g(\lambda)$ 最大，可以改变斜率，多画几个试试

疑问：为什么 p^* 不是蓝点？⇒ p^* 指纵坐标

得出结论: $\boxed{d^* \le p^*}$, 那什么时候有 $d^* = p^*$?



对于凸集, 必然能找到 λ, 使 $g(\lambda) = p^*$

$\boxed{\text{凸优化} + \text{slater条件} \Rightarrow d^* = p^* \text{(强对偶)}}$

- slater 条件

$\exists\, \hat{x} \in \text{relint } D$ relint : relative interior (相对内部)

s.t. $\forall i = 1, \cdots, M,\ m_i(\hat{x}) < 0$

① 对于大多数凸优化问题, slater condition 成立

② 放松的 slater condition : M中有K个仿射函数 (affine), 则只需验证 M-K 个

- KKT条件

Convex + slater $\longrightarrow$ strong Duality

$\Updownarrow$

KKT

primal problem

$$\begin{cases} \min f(x) \\ \text{s.t. } m_i(x) \le 0, \quad i = 1, \cdots, M \\ n_j(x) = 0, \quad j = 1, \cdots, N \end{cases}$$

$$\mathcal{L}(x,\lambda,\eta) = f(x) + \sum_{i=1}^{N}\lambda_i m_i + \sum_{j=1}^{N}\eta_j n_j$$

$$g(\lambda,\eta) = \min_{x} \mathcal{L}(x,\lambda,\eta)$$

Dual Problem.

$$\begin{cases} \max\limits_{\lambda,\eta}\ g(\lambda,\eta) \\ s.t.\ \lambda \geq 0 \end{cases}$$

KKT:

$$\begin{cases} ① \text{ 可行条件} \begin{cases} m_i(x) \leq 0 \\ n_j(x) = 0 \\ \lambda^* \geq 0 \end{cases} \text{(显而易见)} \\ ② \text{ 互补松弛}: \lambda_i m_i = 0 \\ ③ \text{ 梯度为0 } \left.\dfrac{\partial \mathcal{L}(x,\lambda^*,\eta^*)}{\partial x}\right|_{x=x^*} = 0 \end{cases}$$

对于②:

假设原问题在 x^* 时取最优解 p^*,

对偶问题在 λ^*,η^* 时取最优解 d^*

则:

$$\begin{aligned} \underline{d^*} &= \max_{\lambda,\eta} g(\lambda,\eta) \\ &= g(\lambda^*,\eta^*) \\ &= \min_{x} \mathcal{L}(x,\lambda^*,\eta^*) \\ &\leq \mathcal{L}(x^*,\lambda^*,\eta^*) \\ &= f(x^*) + \sum_{i=1}^{M}\lambda_i^* m_i + \sum_{j=1}^{N}\eta_j^* n_j \qquad ① \end{aligned}$$

$\because n_j(x) = 0 \quad \therefore \sum_{j=1}^{N}\eta_j^* n_j = 0$

$\because \begin{cases} m_i(x) \le 0 \\ \lambda^* \ge 0 \end{cases} \quad \therefore \sum_{i=1}^{m} \lambda_i^* m_i \le 0$

$$\therefore ① = f(x^*) + \sum_{i=1}^{m} \lambda_i^* m_i$$
$$\le \underline{f(x^*)} \quad p^*$$

欲使 $d^* = p^*$ (Strong duality)，则需满足 $\underline{\sum_{i=1}^{m} \lambda_i^* m_i = 0}$

即满足 $\boxed{\lambda_i m_i = 0. \ \forall i = 1, \dots, M}$ complementary slack

对于③：

$\because \min_{x} \mathcal{L}(x, \lambda^*, \eta^*) = \mathcal{L}(x^*, \lambda^*, \eta^*)$

$\therefore \boxed{\left.\frac{\partial \mathcal{L}}{\partial x}\right|_{x = x^*} = 0},$

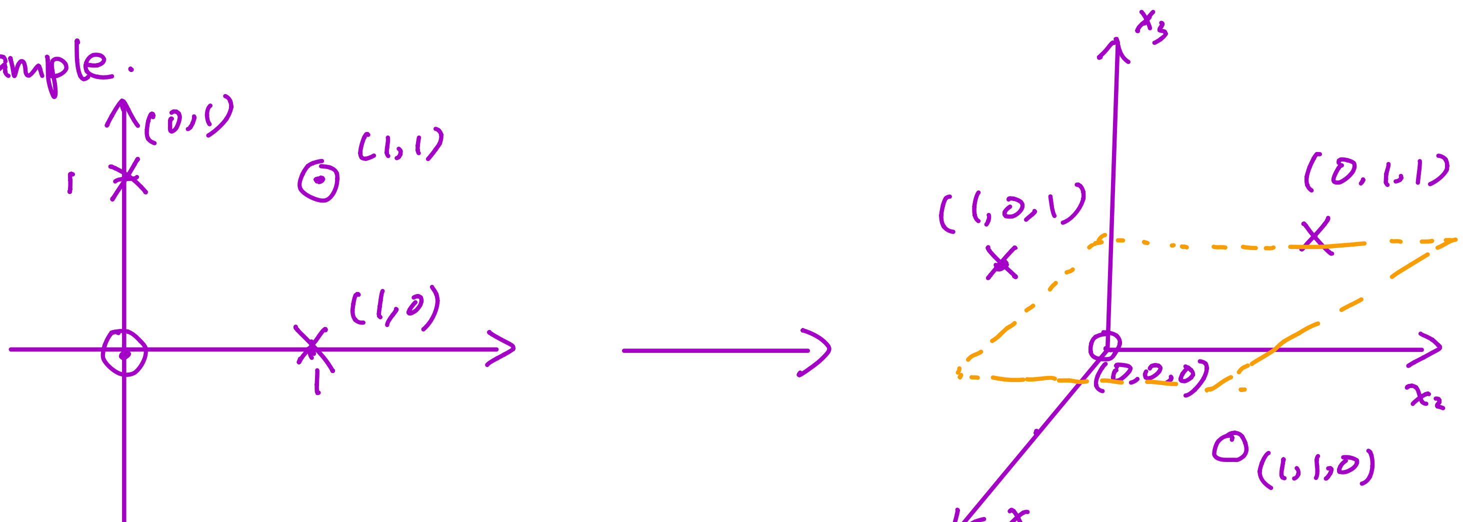
Kernel Method 核方法

- 非线性带来高维转换

$$\mathcal{X} \xrightarrow{\phi(x)} \mathcal{Z}$$

Input space　　　feature space

Example.



异或问题，线性不可分　　　　　　　　线性可分

$x=(x_1,x_2)$ 2d

$\downarrow$

$z=(x_1,x_2,(x_1-x_2)^2)$ 3d

· 对偶表示带来内积

Primal Problem $\longrightarrow$ Dual Problem

$$\begin{cases}\min\limits_{w,b} \frac{1}{2} w^T w \\ s.t \quad y_i(w^T x+b)\geq 1\end{cases}$$

$\phi(x_i)^T\phi(x_j)$

$$\begin{cases}\min \frac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boxed{x_i^T x_j}-\sum\limits_{i=1}^{N}\lambda_i \\ s.t. \quad \lambda_i\geq 0 \quad \forall i=1,2,\cdots,N \\ \qquad \sum\limits_{i=1}^{N}\lambda_i y_i=0\end{cases}$$

在对偶问题中，要求内积 $\phi(x_i)^T\phi(x_j)$，这就需要将样本先投射到高维空间，然后再求在高维空间中的内积，这样会耗费极多的计算资源，为此引入核函数(kernal function)，核函数能事先在低维上进行计算，而将实质的分类效果表现在高维上，避免了在高维上的复杂运算。

定义：$\forall x, x'\in \mathcal{X}$, $\exists\phi: \mathcal{X}\rightarrow \mathcal{Z}$

$s.t.\quad K(x,x')=\phi(x)^T\phi(x')$.

则称 $K(x,x')$ 是一个核函数

Example. $K(x,x')=\exp\left(-\frac{(x-x')^2}{2\sigma^2}\right)$ 是一个核函数.

证明：

$$e^{\left(-\frac{x^2}{2\sigma^2}\right)}\cdot e^{\frac{xx'}{\sigma^2}}\cdot e^{\left(-\frac{x'^2}{2\sigma^2}\right)}$$

$$= e^{\left(-\frac{x^2}{2\sigma^2}\right)}\cdot\sum_{n=0}^{+\infty}\frac{x^n x'^n}{\sigma^{2n}\cdot n!}\, e^{\left(-\frac{x'^2}{2\sigma^2}\right)}$$

$$= e^{\left(-\frac{x^2}{2\sigma^2}\right)}\varphi(x)\cdot\varphi(x')\, e^{\left(-\frac{x'^2}{2\sigma^2}\right)}$$

$$=\phi(x)\cdot\phi(x')$$

· 补充正定核及其性质与证明.

·正定核

定义一: $k: \mathcal{X} \times \mathcal{X} \longmapsto \mathbb{R}$, $\forall x, z \in \mathcal{X}$. 有 $k(x,z)$.

如果 $\exists$: $\phi: \mathcal{X} \leftrightarrow \mathbb{R}$, $\phi \in H$, s.t. $k(x,z) = \langle \phi(x), \phi(z) \rangle$.

那么则称 $k(x,z)$ 为正定核函数.

定义二. $k: \mathcal{X} \times \mathcal{X} \longmapsto \mathbb{R}$, $\forall x, z \in \mathcal{X}$, 有 $k(x,z)$

如果 $k(x,z)$ 满足 $\begin{cases} ① 对称性 \\ ② 正定性 \end{cases}$, 那么称 $k(x,z)$ 为正定核函数.

其中, ① 对称性 $\Longleftrightarrow$ $k(x,z) = k(z,x)$

② 正定性 $\Longleftrightarrow$ 任取N个元素, $x_1, x_2, \cdots, x_N \in \mathcal{X}$, 对应的 Gram matrix $K_{ij} = k(x_i, x_j)$ 是半正定的。

A Matrix $M \in \mathbb{R}^{n \times n}$ is positive semi-definite if $\forall a \in \mathbb{R}^n$, $a' M a \geq 0$

补充 Hilbert space

Hilbert space: 完备的, 可能是无限维的, 被赋予内积的线性空间

① 完备的: 指对极限是封闭的, 即 $\lim_{n \to \infty} k_n = k \in H$.

② 被赋予内积 $\Rightarrow \begin{cases} 对称性 & \langle f, g \rangle = \langle g, f \rangle \\ 正定性 & \langle f, f \rangle \geq 0. \text{ "="} \Longleftrightarrow f = 0 \\ 线性 & \langle r_1 f_1 + r_2 f_2, g \rangle = r_1 \langle f_1, g \rangle + r_2 \langle f_2, g \rangle \end{cases}$

③ 线性空间: 向量空间 (对加法和乘法是封闭的)