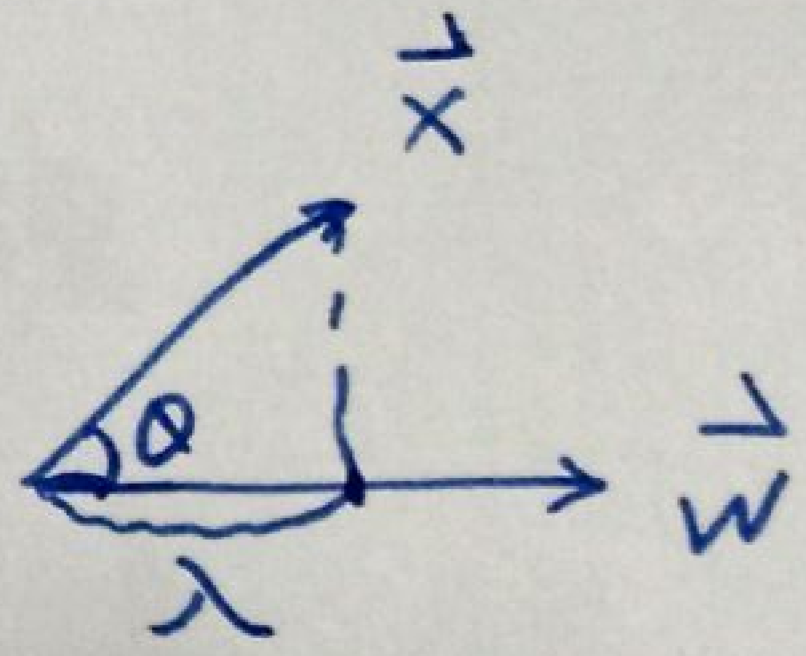
一、内积几何意义

①

$$\vec{w} = (w_1, w_2, \cdots w_n)^T \in R^n$$

$$\vec{x} = (x_1, x_2, \cdots x_n)^T \in R^n$$

$$\langle \vec{w}, \vec{x} \rangle = \vec{w}^T \vec{x} = \sum_i w_i x_i$$



$$= \|\vec{w}\| \|\vec{x}\| \cos\theta$$

$$= \|\vec{w}\| \underbrace{(\|\vec{x}\| \cos\theta)}_{\lambda}$$

令 $\lambda = \|\vec{x}\| \cos\theta$. 为$\vec{x}$向$\vec{w}$方向的投影长度

$$\langle \vec{w}, \vec{x} \rangle = \lambda \|\vec{w}\|$$

② 给定数据 $D=\{\vec{x}_1, \vec{x}_2 \cdots \vec{x}_m\}$

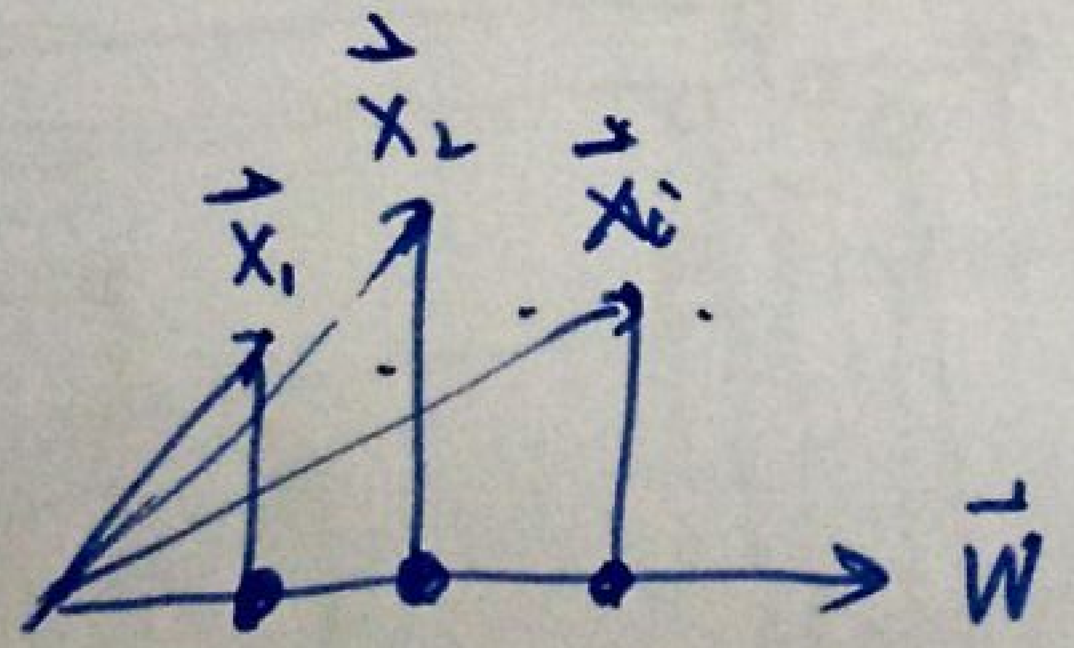
$$\langle \vec{w}, \vec{x}_1 \rangle = \lambda_1 \|W\|$$

$$\langle \vec{w}, \vec{x}_2 \rangle = \lambda_2 \|W\|$$

$$\vdots \qquad \vdots$$

$$\langle \vec{w}, \vec{x}_m \rangle = \lambda_m \|W\|.$$

其中，$\lambda_i = \|\vec{x}_i\| \cos\theta$，为 $\vec{x}_i$ 在 W 方向上投影长度



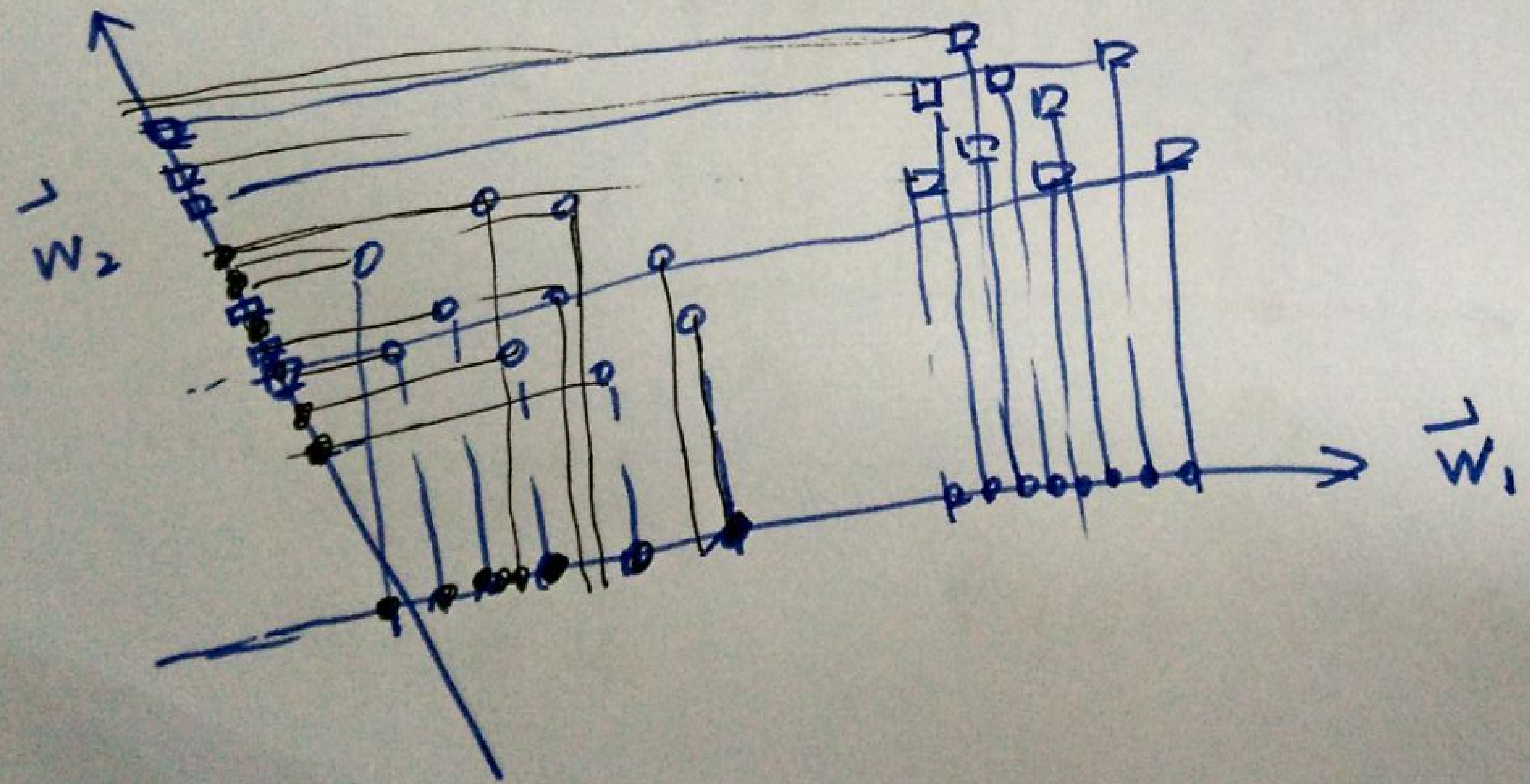
③ $\vec{x}_i \to \lambda_i$，将 $\vec{x}_i$ 降为一维数值 λ_i（忽略 $\|W\|$ 的影响）

二、线性鉴别分析 (Fisher 鉴别)

目标：寻找某一 $\vec{w}$，使得 $D=\{(\vec{x}_i, y_i)\}$ 投影后，

类别内类距离最小，类别之间距离最大。

① 先看不同的 $\vec{w}$ 的效果：两类 o，□。

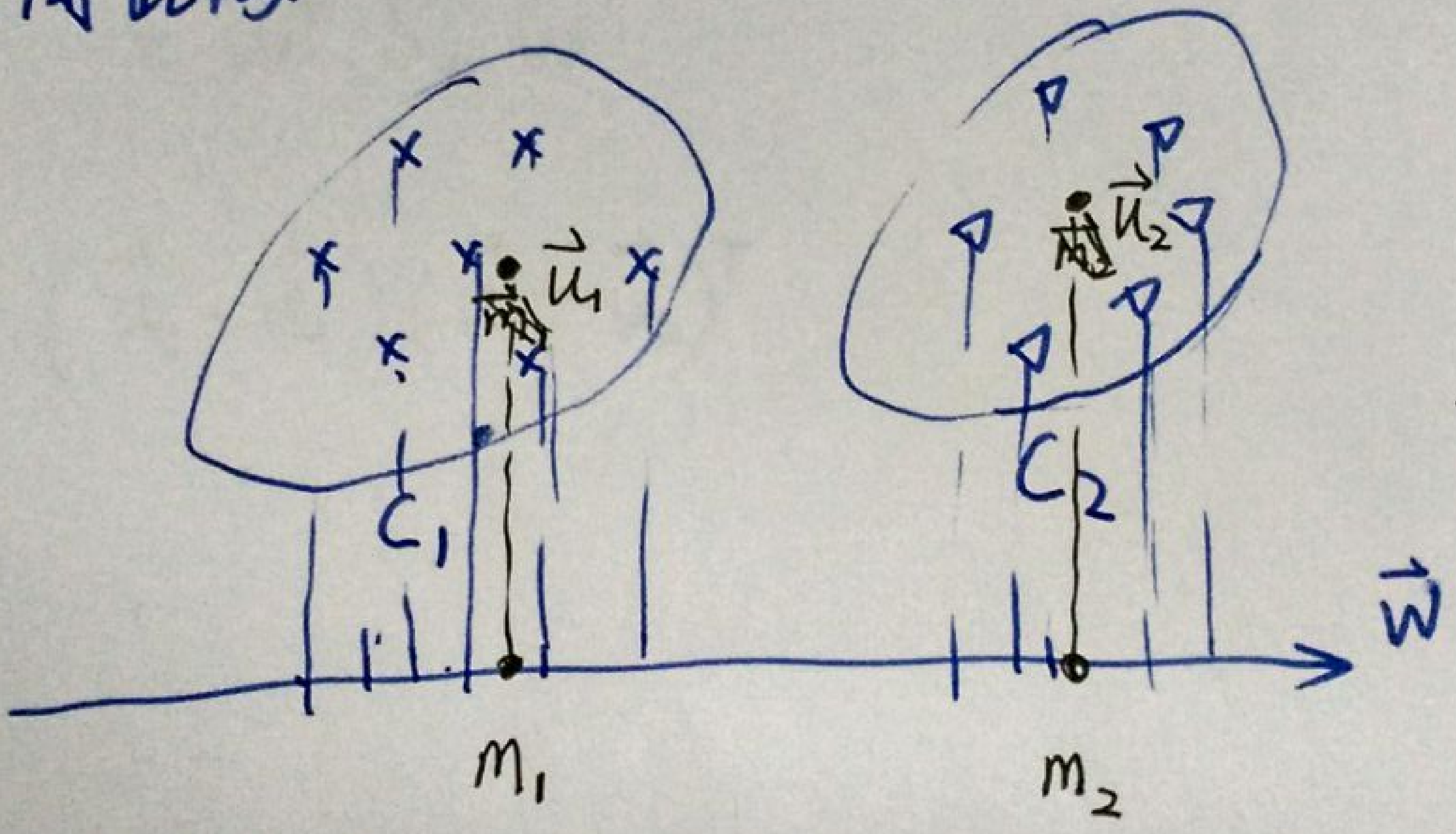


显然 $\vec{w}_1$ 方向比 $\vec{w}_2$ 方向更能使得 $\{\vec{x}_i\}$ 的投影值 $\{x_i\}$ 分的开。

② Fisher 准则

$$\max_{w} J(w) = \frac{\text{类间平均距离}}{\text{类内平均距离}}$$

- 类间距离



$\vec{u}_1$，类别1的样本平均

$$\vec{u}_1 = \frac{1}{N_1}\sum_{i \in C_1} \vec{x}_i$$

$\vec{u}_2$，类别2的样本平均

$$\vec{u}_2 = \frac{1}{N_2}\sum_{i \in C_1} \vec{x}_i$$

$$m_1 = \vec{w}^T \vec{u}_1$$

$$m_2 = \vec{w}^T \vec{u}_2$$

$$\text{类间距离} = (m_1 - m_2)^2$$

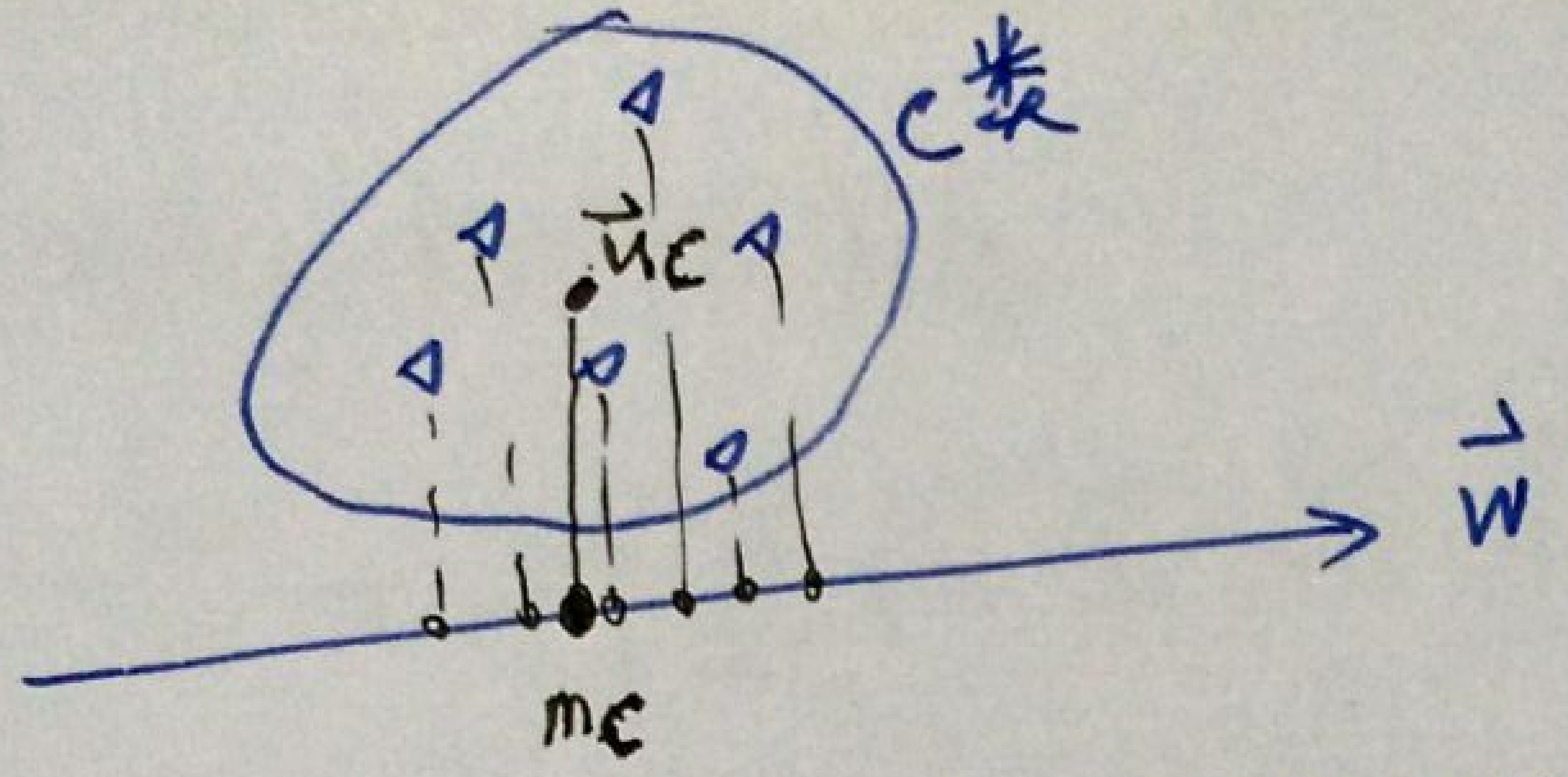
$$(m_1 - m_2)^2 = \left[\vec{w}^T(\vec{u}_1 - \vec{u}_2)\right]\left[\vec{w}^T(\vec{u}_1 - \vec{u}_2)\right]^T$$

$$= \vec{w}^T \underbrace{(\vec{u}_1 - \vec{u}_2)(\vec{u}_1 - \vec{u}_2)^T}_{\text{类间散度矩阵}} \vec{w}$$

$$= \vec{w}^T S_b \vec{w}$$

其中： $S_b = (\vec{u}_1 - \vec{u}_2)(\vec{u}_1 - \vec{u}_2)^T$

• 类内距离



均方差统计类别内部的松散程度，作为类内距离。

$$S_C^2 = \sum_{i\in C} \left\| \vec{w}^T \vec{x}_i - m_C \right\|^2 = \sum_i \left\| \vec{w}^T (\vec{x}_i - \vec{u}_c) \right\|^2$$

$$= \sum_{i\in C} \left[\vec{w}^T(\vec{x}_i - \vec{u}_c)\right]^{\cancel{T}} \left[\vec{w}^T(\vec{x}_i - \vec{u}_c)\right]^T$$

$$= \sum_{i\in C} \vec{w}^T \underbrace{(\vec{x}_i - \vec{u}_c)(\vec{x}_i - \vec{u}_c)^T}_{\text{类内散度矩阵}} \vec{w}$$

$$= \vec{w}^T \left[\sum_{i\in C} (\vec{x}_i - \vec{u}_c)(\vec{x}_i - \vec{u}_c)^T \right] \vec{w}$$

两类问题的类内平均距离

$$S_1^2 + S_2^2 = \vec{w}^T \left(\sum_{i \in C_1} (\vec{x}_i - \vec{u}_1)(\vec{x}_i - \vec{u}_1)^T \right) \vec{w} + \vec{w}^T \left(\sum_{i \in C_2} (\vec{x}_i - \vec{u}_2)(\vec{x}_i - \vec{u}_2)^T \right) \vec{w}$$

$$= \vec{w}^T \underbrace{\left(\sum_{j=1,2} \sum_{i \in C_j} (\vec{x}_i - \vec{u}_j)(\vec{x}_i - \vec{u}_j)^T \right)}_{\text{类内平均散度矩阵}} \vec{w}$$

$$= \vec{w}^T \, S_w \, \vec{w}$$

其中，$S_w = \sum_{j=1,2} \sum_{i \in C_j} (\vec{x}_i - \vec{u}_j)(\vec{x}_i - \vec{u}_j)^T$ ，

目标函数

$$\max_{w} J(w) = \frac{\text{类间平均距离}}{\text{类内平均距离}} = \frac{(m_1 - m_2)^2}{S_1^2 + S_2^2}$$

$$= \frac{\vec{w}^T S_b \vec{w}}{\vec{w}^T S_w \vec{w}}$$

等价于如下问题

$$\max_{w} L(w, \lambda) = \vec{w}^T S_b \vec{w} - \lambda(\vec{w} S_w \vec{w} - c)$$

$$\frac{\partial L(w, \lambda)}{\partial \vec{w}} = S_b \vec{w} - \lambda S_w \vec{w} = 0$$

$$\Rightarrow \quad S_b \vec{w} = \lambda S_w \vec{w}$$

$$\Rightarrow \quad S_w^{-1} S_b \vec{w} = \lambda \vec{w}$$

最优解是 $S_w^{-1} S_b$ 的特征向量

实际没有直接求 · $S_w^{-1} S_b \vec{w} = \lambda \vec{w}$.

因为 $S_b \vec{w}$ 在 $\vec{u}_1 - \vec{u}_2$ 方向上.

$$S_b \vec{w} = \underbrace{(\vec{u}_1 - \vec{u}_2)(\vec{u}_1 - \vec{u}_2)^T}_{S_b} \vec{w} = (\vec{u}_1 - \vec{u}_2) \underbrace{\left[(\vec{u}_1 - \vec{u}_2)^T \vec{w} \right]}_{\text{标量值}}$$

$$= \beta (\vec{u}_1 - \vec{u}_2)$$

代入 $S_w^{-1} S_b \vec{w} = \lambda \vec{w}$

$$S_w^{-1} \underline{\beta} (\vec{u}_1 - \vec{u}_2) = \underline{\lambda} \vec{w} \Rightarrow$$

$\vec{w} = \frac{\beta}{\lambda} S_w^{-1} (\vec{u}_1 - \vec{u}_2)$，方向为尺度 $\frac{\beta}{\lambda}$ 省略.

$$\vec{w} = S_{\vec{w}} (\vec{u}_1 - \vec{u}_2)$$