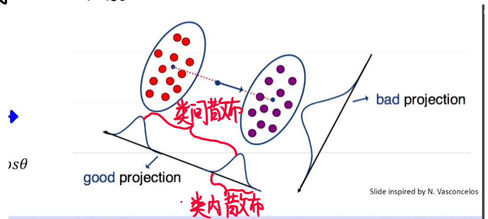
§4.1 Fisher线性判别动机

人图像是高维特征空间中的一个点，一类图像在空间中分布相对集中

* 核心思想：捕获人脸关键特征并压缩到低维空间

将 x 向 w 投影 $w^T x = \|x\|\|w\|\cos\theta$



Fisher线性判别的目的：在两个类别之间找到最好的区分
(*find the best separation between two classes*)

Fisher线性判别目的

①尽可能保留类别可区分性前提下实现维数减少

②找到让类别最好区分的投影方向

③同时考虑类内散布和类间散布

§4.2 Fisher线性判别分析

二分类问题的Fisher线性判别

学习最佳投影 w^*，将所有样本投影到 w^* 方向

假设 $s = \boldsymbol{w}^T \mathbf{x}$ $\quad \mathbf{x} \in \mathcal{R}^d, \ s \in \mathcal{R}^1$

类别集合：$\mathcal{C} = \{c|(1,-1)\}$

第 c 个类别的均值为：$\boldsymbol{\mu}_c = E[\mathbf{x}|y=c] = \frac{1}{N_c}\sum_{n=1}^{N_c}[\mathbf{x}_n|y=c]$

第 c 个类别的协方差为：$\boldsymbol{\Sigma}_c = E[(\mathbf{x}-\mu_c)(\mathbf{x}-\mu_c)^T|y=c]$
$= \sum_{n=1}^{N_c}[(\mathbf{x}_n-\mu_c)(\mathbf{x}_n-\mu_c)^T|y=c]$

学习最佳投影

$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{(E[s|y=1]-E[s|y=-1])^2}{var[s|y=1]+var[s|y=-1]}$$

$$\begin{aligned}&(E[s|y=1]-E[s|y=-1])^2\\ &= (E[\boldsymbol{w}^T\mathbf{x}|y=1]-E[\boldsymbol{w}^T\mathbf{x}|y=-1])^2\\ &= (\boldsymbol{w}^T(E[\mathbf{x}|y=1]-E[\mathbf{x}|y=-1]))^2\\ &= (\boldsymbol{w}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1}))^2\\ &= \boldsymbol{w}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})^T\boldsymbol{w}\end{aligned}$$

$$\begin{aligned}var[s|y=c] &= E[(s-E[s|y=c])^2]\\ &= E\left[\left(\boldsymbol{w}^T\mathbf{x}-E[\boldsymbol{w}^T\mathbf{x}|y=c]\right)^2\right]\\ &= E\left[\left(\boldsymbol{w}^T(\mathbf{x}-E[\mathbf{x}|y=c])\right)^2\right]\\ &= E\left[\left(\boldsymbol{w}^T(\mathbf{x}-\boldsymbol{\mu}_c)\right)^2\right]\\ &= E[\boldsymbol{w}^T(\mathbf{x}-\boldsymbol{\mu}_c)(\mathbf{x}-\boldsymbol{\mu}_c)^T\boldsymbol{w}]\\ &= \boldsymbol{w}^T E[(\mathbf{x}-\boldsymbol{\mu}_c)(\mathbf{x}-\boldsymbol{\mu}_c)^T]\boldsymbol{w}\\ &= \boldsymbol{w}^T\boldsymbol{\Sigma}_c\boldsymbol{w}\end{aligned}$$

19

$\Rightarrow J(w) = \dfrac{w^T(\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T w}{w^T\Sigma_1 w + w^T\Sigma_{-1} w}$，记 $S_B = (\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T$，$S_W = \Sigma_1+\Sigma_{-1} = S_1+S_2$

$\Rightarrow J(w) = \dfrac{w^T S_B w}{w^T S_W w}$

最大化目标函数问题转化为

$$\max_{w} \ w^T S_B w$$
$$s.t. \ w^T S_W w = k$$

Lagrange multipliers 法：

$L(w,\lambda) = w^T S_B w - \lambda(w^T S_W w - k) = w^T(S_B - \lambda S_W)w - \lambda k$

令 $\nabla L_w(w,\lambda) = \dfrac{\partial L(w,\lambda)}{\partial w} = 0 \Rightarrow 2(S_B-\lambda S_W)w = 0 \Rightarrow S_B w = \lambda S_W w$

若 $S_W^{-1} = (\Sigma_1+\Sigma_{-1})^{-1}$ 存在，则有

$$S_W^{-1}S_B w = \lambda w$$

$$S_W^{-1}(\mu_1-\mu_{-1})\underbrace{(\mu_1-\mu_{-1})^T w}_{a} = \lambda w$$

$$\frac{\lambda}{a} w = S_W^{-1}(\mu_1-\mu_{-1})$$

所以 $w^* = S_W^{-1}(\mu_1-\mu_{-1})$

找到投影向量后，对任一测试样本 $\mathbf{x}$：

$$s = \boldsymbol{w}^{*T}\mathbf{x} = (S_W^{-1}(\mu_1-\mu_{-1}))^T\mathbf{x}$$

假设类别的判别门限设为 s'：

$$s' = \frac{\boldsymbol{w}^{*T}(\mu_1+\mu_{-1})}{2}$$

对任一测试样本 $\mathbf{x}$ 所属类别的判断：

$$\begin{cases} y = 1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} > s' \\ y = -1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} < s' \end{cases}$$

§4.3 Fisher线性判别算法

① 获取具有标签的两类样本

② 依据下式得到 $\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_{-1}$：

$$\boldsymbol{\mu}_c = \frac{1}{N_c}\sum_{n=1}^{N_c}[\mathbf{x}_n|y=c]$$

③ 依据下式得到 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_{-1}$：

$$\boldsymbol{\Sigma}_c = \sum_{n=1}^{N_c}[(\mathbf{x}_n-\mu_c)(\mathbf{x}_n-\mu_c)^T|y=c]$$

④ 计算类内总离差阵：$\boldsymbol{S}_w = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{-1}$

协方差矩阵和

⑤ 计算类内总离差阵的逆：$\boldsymbol{S}_w^{-1}$

⑥ 计算最佳投影：$\boldsymbol{w}^* = \boldsymbol{S}_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})$

⑦ 计算判别门限 s'：$s' = \dfrac{\boldsymbol{w}^{*T}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_{-1})}{2}$

⑧ 对任一测试样本 $\mathbf{x}$：

$$\begin{cases} y = 1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} > s' \\ y = -1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} < s' \end{cases}$$