x_{ij} expression, gene i, sample j $i: 1 \sim p$ $j: 1 \sim n$.

class: $1 \sim K$.

C_k k-th class, which has n_k samples. C_k是索引.

$\bar{x}_{ik} = \sum_{j \in C_k} \frac{x_{ij}}{n_k}$ gene i of the centroid for class k.

$\bar{x}_i = \sum_{j=1}^{n} \frac{x_{ij}}{n}$ overall centroid, gene i.

shrink class centroid $\bar{x}_{ik} \rightarrow \bar{x}_i$

Standardization for gene i.

$d_{ik} = \frac{\bar{x}_{ik} - \bar{x}_i}{m_k (s_i + s_0)}$ $\bar{x}_{ik} - \bar{x}_i$: between class.

$s_i: = \frac{1}{n-K} \sum_k \sum_{j \in C_k} (x_{ij} - \bar{x}_{ik})^2$: within class (pooled: sum of all class)

within class

pooled.

$m_k = \sqrt{\frac{1}{n_k} + \frac{1}{n}}$. 有. $m_k \cdot s_i \sim (\bar{x}_{ik} - \bar{x}_i)$'s standard error.

s_0 positive constant (for all genes)

防止 d_{ik} 过大 (来自 lowexpression level's genes).

取 $s_0 = \underset{s_i}{\text{median}}(s_i)$.

rewrite: $\bar{x}_{ik} = \bar{x}_i + m_k (s_i - s_0) d_{ik}$.

Solution: shrink each d_{ik} foward 0 (why?).

$\bar{x}'_{ik} = \bar{x}_i + m_k (s_i + s_0) d'_{ik}$

We call this soft thresholding. $d'_{ik} = \text{sign}(d_{ik}) (|d_{ik}| - \Delta)_+$. $\begin{cases} t_+ = t & t > 0 \\ t_+ = 0 & t \le 0 \end{cases}$

when $\Delta \uparrow$, more gene would be eliminated from the class.

for a gene i, $d_{ik} \xrightarrow{\Delta} 0$, the centroid of gene i is $\bar{x}_i$, that means gene i does not contrite to the nearest-centroid computation.

1st. Get gene expression profile of each test sample

2nd. 计算到每个类重心的平方距离.

判别项 for class k.

$$\delta_k(x^*) = \sum_{i=1}^{p} \frac{(x_i^* - \bar{x}'_{ik})^2}{(s_i + s_0)^2} - 2\log \pi_k.$$

$$\begin{cases} \bar{x}'_{ik} = \bar{x}_i + m_k(s_i + s_0)\, d'_{ik} \\ d'_{ik} = \mathrm{sign}(d_{ik})\,(|d_{ik}| - \Delta)_+ \end{cases}$$

π_k : class prior prop probability.

$\sum_{k=1}^{K} \pi_k = 1$. $\simeq$ $\hat{\pi}_k = \frac{n_k}{n}$. or $\hat{\pi}_k = \frac{1}{K}$

~~Then $C(x^*) = *$~~

$C(x^*) = l$ where $\delta_l(x^*) = \min_k(\delta_k(x^*))$. $l \in 1 \sim k$.

类概率 for estimates.

$$\hat{P}_k(x^*) = \frac{e^{-\frac{1}{2}\delta_k(x^*)}}{\sum_{l=1}^{K} e^{-\frac{1}{2}\delta_l(x^*)}}$$

Appendix.

~~$\delta_k^{LDA}(x^*) = (x^* - \bar{x}'_k)\, W^{-1}\, (x^* - \bar{x}'_k) - 2\log \pi_k$~~

$$\delta_k^{LDA}(x^*) = (x^* - \bar{x}_k)\, W^{*-1}\, (x^* - \bar{x}_k) - 2\log \pi_k.$$