Gaussian Discriminant Analysis

Data: $y \sim$ discrete nominal data labeling classes

$x \sim$ features with which we hope to infer y using $p(y|x)$

Generative method: Model $p(x,y)$ or $p(x|y)$

$$\longrightarrow \quad p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Model: $y \sim \text{Multinomial}(n=1, \text{pvals} = [p_1, p_2, 1-p_1-p_2])$

$$x|(y=i) \sim \mathcal{N}(\vec{\mu} = \vec{\mu}_i, \Sigma = \Sigma_i)$$

Maximum probability Classification

* Use max log likelihood to find estimator of parameters
  - likelihood of JOINT DISTRIBUTION.
    - $p(x,y) = p(x|y)p(y)$ & $p(x,y) = p(y|x)p(x)$
  - apparently there always exist closed-form solutions.

Minimum Distance Classification

* Alternative: Following Rencher (or Adachi) for two-group discrimination,

take $y_1^{(i)}$ & $y_2^{(i)} \sim$ samples of pops 1 & 2.

$z_k^{(i)} = a \cdot y_k^{(i)}$. Minimize $\dfrac{(\bar{z}_1 - \bar{z}_2)^2}{a^T S_{pooled}\, a}$

where $S_{pooled} = \dfrac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} \sim$ smaller variance

estimator of shared covariance, $S_i \sim$ sample covariance matrix.

$$\Rightarrow a = S_{pooled}^{-1}(\bar{y}_1 - \bar{y}_2) \sim \text{Linear Discriminant}$$

Max Likelihood Approach

sets of parameters

$$\ell(p_i, \mu_i, \Sigma_i) = \log \prod_{k=1}^{N} p(x^{(k)}, y^{(k)}; p_i, \mu_i, \Sigma_i)$$

$$= \sum_{k=1}^{N} \log\left[p(x^{(k)} | y^{(k)}; \mu_i, \Sigma_i) \, p(y^{(i)}; p_i) \right]$$

$$\ell(p_i, \mu_i, \Sigma_i) = \sum_{k=1}^{N} \left[\log p(y^{(k)}; p_i) + \log p(x^{(k)} | y^{(k)}; \mu_i, \Sigma_i) \right]$$

First : Take $i = 0, 1$: $p(y^{(k)}; p_i) = p_i^{y^{(k)}} (1 - p_i)^{1 - y^{(k)}}$

$$\partial_{p_1} \ell = \sum_{k=1}^{N} \partial_{p_1} \left[y^{(k)} \log p_1 + (1 - y^{(k)}) \log(1 - p_1) \right] = 0$$

$$= \sum_{k=1}^{N} \left[1\{y^{(k)} = 1\} \frac{1}{p_1} - 1\{y^{(k)} = 0\} \frac{1}{1 - p_1} \right] = 0$$

$$(1 - p_1) \sum_k 1\{y^{(k)} = 1\} = p_1 \sum 1\{y^{(k)} = 0\} = p_1 (N - n_1)$$

$$(1 - p_1) n_1 = p_1 (N - n_1) \implies n_1 - n_1 p_1 = N p_1 - n_1 p_1$$

$$\implies p_1 = \frac{n_1}{N} \implies \boxed{p_c = n_c / N}$$

Check Guess Below :

$$\partial_{p_c} \ell = \sum_k \partial_{p_c} \log(y^{(k)}; p_i = \{p_0, p_1, \ldots, 1 - (p_0 + p_1 \cdots p_{c-1})\})$$

$$0 = \sum_k \left[\frac{1\{y^{(k)}=c\}}{p_c} - \frac{1\{y^{(k)}=L\}}{1-(p_0+p_1+\dots p_{L-1})} \right]$$

$$(1-(p_0+p_1+\dots+p_c+\dots+p_{L-1}))\,n_c = p_c n_L = p_c(N-(n_0+n_1+\overset{n_c}{\cdots}+n_{L-1}))$$

$$-p_c n_c + (1-(p_0+p_1+\dots 0+\dots+p_{L-1}))n_c = -n_c p_c + p_c(N-(n_0+n_1+\overset{\cancel{n_c}}{\cdots}+n_{L-1}))$$

$$\left(1-\sum_{i\neq c}^{L-1} p_i\right)n_c = p_c\left(N-\sum_{i\neq c}^{L-1} n_i\right)$$

$$\boxed{\text{Guess: } p_i = \frac{n_i}{N}}$$

$$\left(1-\frac{1}{N}\sum_{i\neq c}^{L-1} n_i\right)n_c = \frac{n_c}{N}\left(N-\sum_{i\neq c}^{L-1} n_i\right)$$

$$\left(N-\sum_{i\neq c}^{L-1} n_i\right)n_c = n_c\left(N-\sum_{i\neq c}^{L-1} n_i\right)$$

Finding Means

$$\partial_{\mu_c} l = \sum_{k=1}^{N} \partial_{\mu_c} \sum_{i=1}^{L}\left[-\log|\Sigma_i|^{1/2} - \frac{1}{2}(x^{(k)}-\mu_i)^T\Sigma_i^{-1}(x^{(k)}-\mu_i)\right]$$

$$0 = \sum_{k=1}^{N} 1\{y^{(k)}=c\}\,\Sigma_c^{-1}(x^{(k)}-\mu_c) = \Sigma_c^{-1}\sum_{k=1}^{N}(x^{(k)}-\mu_c)\,1\{y^{(k)}=c\}$$

$$\Longrightarrow \sum_{k=1}^{N}(x^{(k)}-\mu_c)\,1\{y^{(k)}=c\} = 0$$

$$\Longrightarrow \mu_c = \frac{\sum 1\{y^{(k)}=c\}\,x^{(k)}}{n_c}$$

Finding Variances

$$-\frac{1}{2}\operatorname{Trace}\log\Sigma_c$$

$$\partial_{\Sigma_c} \ell = \sum_{k=1}^{N} 1\{y^{(k)}=c\}\, \partial_{\Sigma_c}\Big[-\frac{1}{2}\log|\Sigma_c^{-1}|$$

$$-\frac{1}{2}(x^{(k)}-\mu_c)^T \Sigma_c^{-1}(x^{(k)}-\mu_c)\Big]$$

Using $\partial_X \operatorname{tr}\log X = X^{-1}$; $\partial_{\Sigma^{-1}} \operatorname{tr}\log \Sigma^{-1} = \Sigma^{-1^T} = \Sigma^{-1}$

Adachi 4.5.2 :

$$(x^{(k)}-\mu_c)^T\Sigma_c^{-1}(x^{(k)}-\mu_c) = \operatorname{Tr}\Sigma_c^{-1}(x^{(k)}-\mu_c)(x^{(k)}-\mu_c)^T$$

$$= \operatorname{Tr}\Sigma_c^{-1}A$$

Wikepedia (Matrix Calculus) : ① $\partial_X \operatorname{tr}(X^{-1}A) = -X^{-1}AX^{-1}$

$\partial_X \operatorname{tr}(XA) = A$

$\partial_X a^T X b = b a^T$

$\partial_X \operatorname{tr}(g(X)) = g'(X)$

Proof ①: $d(XX^{-1}) = 0 \quad dX\, X^{-1} + X\, dX^{-1} = 0$

$$dX^{-1} = -X^{-1}\,dX\,X^{-1}$$

$$\Rightarrow \partial_X F = -X^{-1}\,\partial_X F\, X^{-1}$$

$$\partial_{\Sigma_c} \operatorname{Tr}\Sigma_c^{-1}A = -\Sigma_c^{-1}A\Sigma_c^{-1}$$

$$\partial_{\Sigma_c^{-1}} \ell = \sum_{k=1}^{N} 1\{y^{(k)}=c\}\Big[-\frac{1}{2}\Sigma_c^{-1} + \frac{1}{2}\Sigma_c^{-1}A\Sigma_c^{-1}\Big] = 0$$

$$\sum_{k=1}^{N} 1\{y^{(k)}=c\}\,\Sigma_c^{-1} = \sum_{k=1}^{N} 1\{y^{(k)}=c\}\,\Sigma_c^{-1}A\Sigma_c^{-1}$$

$$\sum_{k=1}^{N} 1\{y^{(k)}=c\}\,\Sigma^{-1}_c = \sum_{k=1}^{N} 1\{y^{(k)}=c\}\,A$$

$$\Sigma^{-1}_c = \frac{\sum_{k=1}^{N} 1\{y^{(k)}=c\}(x^{(k)}-\mu_c)(x^{(k)}-\mu_c)^T}{n_c = \sum_{k=1}^{N} 1\{y^{(k)}=c\}}$$

Finding the Decision Boundary

Determine lines along which $p(y=i\,|\,x) = p(y=j\,|\,x) = \dots$

where # classes $= L$. Use $p(y\,|\,x) = \frac{p(x|y)\,p(y)}{p(x)}$ ← cancel in equations

Easier: $\log p(y=i\,|\,x) = \log p(y=j\,|\,x)$ ~ boundary $f_{ij}(x)=0$

Merging lines: $\log p(y=i|x) = \frac{1}{2}$ → intersection point (should be same $\forall i$?)

- all lines $f_{ij}(x)$ intersect at the above point?
  - why not an area w/ equal probability $= 1/L$?

$$\log p_i - \tfrac{1}{2}\operatorname{tr}\log\Sigma^{-1}_i - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}_i(x-\mu_i)(x-\mu_i)^T =$$

$$\log p_j - \tfrac{1}{2}\operatorname{tr}\log\Sigma^{-1}_j - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}_j(x-\mu_j)(x-\mu_j)^T$$

Note: $\Sigma^{-1}_i = \Sigma^{-1}_j \Rightarrow$ linear equation (see below for two group case)

$$\log p_i/p_j + \tfrac{1}{2}\operatorname{tr}\left(\log\Sigma^{-1}_j - \log\Sigma^{-1}_i\right) + \tfrac{1}{2}\operatorname{tr}\left[\Sigma^{-1}_j(x-\mu_j)(x-\mu_j)^T - \Sigma^{-1}_i(x-\mu_i)(x-\mu_i)^T\right] = 0$$

$$\tfrac{1}{2}\left[x^T(\Sigma^{-1}_j - \Sigma^{-1}_i)x - 2x^T(\Sigma^{-1}_j\mu_j - \Sigma^{-1}_i\mu_i) + (\mu_j^T\Sigma^{-1}_j\mu_j - \mu_i^T\Sigma^{-1}_i\mu_i)\right]$$

$$= -\log p_i/p_j - \tfrac{1}{2}\,\mathrm{tr}\left(\log \Sigma_i^{-1} - \Sigma_j^{-1}\right)$$

$$\boxed{x^T(\Sigma_j^{-1} - \Sigma_i^{-1})x - 2x^T(\Sigma_j^{-1}\mu_j - \Sigma_i^{-1}\mu_i) = 2\log p_j/p_i - \mathrm{tr}\left(\log \Sigma_j^{-1} - \log \Sigma_i^{-1}\right) + 2\left(\mu_j^T \Sigma_j \mu_j - \mu_i^T \Sigma_i^{-1} \mu_i\right)}$$

~ Quadratic discriminant

In code: easier to do exhaustive contouring whereby one classifies all cells of the space which are sufficiently small.

Only determining for two groups

$$\log(p_1) - \frac{1}{2}(\vec{x}-\vec{\mu}_1)\Sigma^{-1}(\vec{x}-\vec{\mu}_1) = \log p_0 - \frac{1}{2}(\vec{x}-\vec{\mu}_0)\Sigma^{-1}(\vec{x}-\vec{\mu}_0)$$

$$\log\left(\frac{p_1}{p_0}\right) = \frac{1}{2}(\vec{x}-\vec{\mu}_1)\Sigma^{-1}(\vec{x}-\vec{\mu}_1) - \frac{1}{2}(\vec{x}-\vec{\mu}_0)\Sigma^{-1}(\vec{x}-\vec{\mu}_0)$$

$$2\log\left(\frac{p_1}{p_0}\right) = \vec{x}\Sigma^{-1}\vec{x} - \vec{x}\Sigma^{-1}\mu_1 - \mu_1\Sigma^{-1}x + \vec{\mu}_1\Sigma^{-1}\vec{\mu}_1$$

$$- \left(x\Sigma^{-1}x - \vec{x}\Sigma^{-1}\mu_0 - \mu_0\Sigma^{-1}x + \vec{\mu}_0\Sigma^{-1}\mu_0\right)$$

$$= \vec{x}\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1) + (\vec{\mu}_0-\vec{\mu}_1)\Sigma^{-1}\vec{x} + \mu_1\Sigma^{-1}\mu_1 - \mu_0\Sigma^{-1}\mu_0$$

$$= 2\vec{x}\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1) - \mu_0^2 + \mu_1^2$$

$$\mu_i^2 = \mu_i\Sigma^{-1}\mu_i$$

$$\vec{x}\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1) = \log\left(\frac{p_1}{p_0}\right) - \frac{1}{2}\left(-\mu_0^2 + \mu_1^2\right)$$

$\Rightarrow$ set of linear equations on $\{x_i\}$.