Resumen AED.

1. PCA. Buscamos transformar p variables en q combinaciones lineales ortogonal $x' = (x_1, x_2, \dots, x_p)$ $\quad \lambda_1 \geq \dots \geq \lambda_p \geq 0$. El i-esimo componente principal:

$$y_i = e_i' X = e_{i1} x_1 + e_{i2} x_2 + \dots + e_{ip} x_p \qquad var(y_i) = \lambda_i \qquad cov(y_i, y_k) = 0$$

En general, $\quad y = P'x \quad cov(y) = \Sigma_y = P' \Sigma_x P$

$$\Sigma_x = P \Lambda P', \quad \text{luego} \quad \Sigma_y = diag(\lambda_i)$$

luego, $\sigma_{11} + \dots + \sigma_{pp} = \sum \sigma_{ii} = \lambda_1 + \dots + \lambda_p = \sum \lambda_i = var(y_i)$

→ Proporción total de varianza: $prop(y_k) = \lambda_k / \sum \lambda_i$

→ Correlación: $\rho_{y_i, x_k} = \dfrac{cov(y_i, x_k)}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \dfrac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$

* Para interpretación sacar ρ y Diap 22 para ejemplo

1.1 Población multivariada Normal $\quad X \sim N_p(\mu, \Sigma)$

↳ Elipsoide con centro en μ y con ejes $\mu \pm c\sqrt{\lambda_i}\, e_i$

- Cuando las varianzas son muy diferentes, entonces estandarizamos

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad Z = V^{-1/2}(X - \mu), \quad V^{-1/2} = diag\left(\frac{1}{\sqrt{\sigma_{ii}}}\right)$$

$$\mathbb{E}(Z) = 0 \text{ y } \Sigma_Z = R \Rightarrow \tilde{y} = \tilde{e}_i' Z, \quad \sum var(Z_i) = tr(R) = p$$

Ej: Men's track data, diapositiva 32

1.2 Sample principal components. S estimador de Σ, $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$.

$\boxed{\text{Si tiene } \hat{\ } \text{ es de } S}$

$$\hat{y}_i = \hat{e}_i X = \hat{e}_{i1} x_1 + \dots + \hat{e}_{ip} x_p, \quad var(\hat{y}_i) = \hat{\lambda}_i, \quad cov(\hat{y}_i, \hat{y}_k) = 0$$

- $tr(S) = \sum s_{ii} = \sum \hat{\lambda}_i$ $\quad$ prop-var: $\dfrac{\hat{\lambda}_i}{\sum \hat{\lambda}_k}$

correlación: $r_{\hat{y}_i, x_k} = \dfrac{\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}} \hat{e}_{ik}$ $\quad$ • Si x's son estandar, $r_{\hat{y}_i, x_k} = \sqrt{\hat{\lambda}_i}\, \hat{e}_{ik}$

* las componentes principales son el vector direccional del elipsoide.

Ej. Otro ejemplo de interpretación, diapositiva 57.

2. FA: Buscamos un modelo. $X \to$ obs con μ y Σ. Factores comunes F_i. Error ε_i

↳ Modelo: $\underset{(p\times 1)}{X - \mu} = \underset{(p\times m)}{L} \times \underset{(m\times 1)}{F} + \underset{(p\times 1)}{\varepsilon}$ $\qquad E(F) = 0 \quad cov(F) = I$

$$E(\varepsilon) = 0 \qquad cov(\varepsilon) = \psi = \begin{bmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_p \end{bmatrix} \qquad cov(F, \varepsilon) = 0$$

$$cov(X) = \Sigma = LL' + \psi \qquad cov(X, F) = L$$

$var(X_i) = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$, $\quad cov(X_i, X_k) = l_{i1} l_{k1} + \dots + l_{im} l_{km}$, $\quad cov(X_i, F_j) = l_{ij}$

- Comunalidad: $h_i^2 = \sum l_{ik}^2$, $\quad \sigma_{ii} = h_i^2 + \psi_i$

* Si $m = p$, entonces $\Sigma = LL'$, es decir, ψ es nula.

2.1 Rotación: Sea T una matriz $m \times m$ ortogonal. $\quad X - \mu = L^* F^* + \varepsilon$

$$L^* = LT \qquad F^* = T'F \qquad E(F^*) = 0 \qquad cov(F^*) = I$$

En general, $F \approx F^*$ $\quad L \neq L^*$ $\quad \Sigma = (L^*)(L^*)' + \psi$.

2.2 Metodo de Estimación $\rightarrow$ PCA. Eigen-descomposición de S.
$\rightarrow$ MLE. estimador de maxima verosimilitud.

2.2.1 PCA. Sacamos autovalores y autovectores de S. $\lambda_1 \geq \ldots \geq \lambda_p$

$\tilde{L} = [\sqrt{\hat{\lambda}_1}\hat{e}_1 \ \ldots \ \sqrt{\hat{\lambda}_m}\hat{e}_m]$, $\tilde{\psi} = S - \tilde{L}\tilde{L}'$, $\tilde{\psi}_i = S_{ii} - \sum \tilde{l}_{ij}^2 = S_{ii} - h_i^2$

- $SS(S - (\tilde{L}\tilde{L}' + \tilde{\psi})) \leq \hat{\lambda}_{m+1}^2 + \ldots + \hat{\lambda}_p^2$, $SS(A) = tr(AA')$

En general, proporción total de la varianza para factor j: $\begin{cases} \dfrac{\hat{\lambda}_j}{S_{11} + \ldots + S_{pp}} & \text{para } S \\ \dfrac{\hat{\lambda}_j}{p} & \text{para } R \end{cases}$

Estimación de factores:

Algoritmo:
1. Obtener estimación inicial de $\psi \rightarrow$ SMC de X_i y las otras. El SMC es la diagonal de $R^{-1} \rightarrow \psi^*$
2. Encontrar λ_q^* y e_i^* de $R - diag(\psi_i^*)$. luego. $L^* = (\sqrt{\lambda_1^*}e_1^*, \ldots, \sqrt{\lambda_m^*}e_m^*)$
3. $\psi_i^* = 1 - h_i^{*2} = 1 - \sum l_{iq}^{*2}$
4. Repetir 2 y 3 hasta obtener convergencia.

2.2.2 MLE. Obtenemos el estimador de maxima verosimilitud

$$L(\mu, \Sigma) = (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} \exp\left[-\tfrac{1}{2} tr\left(\Sigma^{-1}\left(\sum (x_j - \bar{x})(x_j - \bar{x})'\right)\right)\right]$$
$$\times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-1/2} e^{-n/2(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)}$$

Proporción total de la varianza para factor $j \rightarrow \dfrac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \ldots + \hat{l}_{pj}^2}{S_{11} + S_{22} + \ldots + S_{pp}}$

Con $Z = V^{-1/2}(X - \mu)$ $\quad \rho = V^{-1/2}\Sigma V^{-1/2}$ $\quad \psi_z = V^{-1/2}\psi V^{-1/2}$

$L_z = V^{-1/2}L$ $\quad \hat{\rho} = \hat{L}_z\hat{L}_z' + \hat{\psi}_z$ $\quad \hat{h}_i^2 = \hat{l}_{i1}^2 + \ldots + \hat{l}_{im}^2$

proporcion total de la varianza para factor $j \rightarrow \dfrac{\hat{l}_{1j}^2 + \ldots + l_{pj}^2}{p}$

Factores: $\hat{L} = \hat{V}^{1/2}\hat{L}_z$ $\quad \hat{\psi} = \hat{V}^{1/2}\hat{\psi}_z\hat{V}^{1/2}$

2.3. Rotación: Hay 2 tipos ortogonal y oblicua.

2.3.1 Ortogonal
T matriz ortogonal, $\Sigma = L^*L^{*\prime} + \psi$, $L^* = LT$ con las estimaciones es igual

2.4. Factor Scores:
$\hat{f}_j^* = T'\hat{f}_j$ $\quad \hat{f}_j = \hat{L}'S^{-1}(x_j - \bar{x}) = \hat{L}_z'R^{-1}z_j$

$z_j = D^{-1/2}(x_j - \bar{x})$ $\quad \hat{\rho} = \hat{L}_z\hat{L}_z' + \hat{\psi}_z$

3. Clasificación y Discriminantes.

3.1 Clasificación 2 poblaciones
$X' = [X_1, \ldots, X_p]$ $\quad f_1(x)$ y $f_2(x)$. π_1 y π_2

Si $x_0 \in R_1 \Rightarrow x_0 \in \pi_1$. Si $x_0 \in R_2 \Rightarrow x_0 \in \pi_2$

$P(2|1) = \int_{R_2} f_1(x)\,dx$ $\quad P(1|2) = \int_{R_1} f_2(x)\,dx$

Costo de clasificación:

	π_1	π_2
π_1	0	$C(2\|1)$
π_2	$C(1\|2)$	0

clasificación, pred $\rightarrow$ columnas
verdadero $\rightarrow$ filas.

→ valor esperado del costo: $ECM = C(2|1)P(2|1)p_1 + C(1|2)P(1|2)p_2$

* la idea es minimizar el costo, una buena clasificación tiene ECM pequeño

$R_1: \dfrac{f_1(x)}{f_2(x)} \geq \dfrac{C(1|2)}{C(2|1)} \cdot \dfrac{p_2}{p_1}$ $\quad R_2: \dfrac{f_1(x)}{f_2(x)} < \dfrac{C(1|2)}{C(2|1)} \cdot \dfrac{p_2}{p_1}$

→ Bayes $\dfrac{p_1 f_1(x)}{p_1 f_1(x) + p_2 f_2(x)} \geq \dfrac{p_2 f_2(x)}{p_1 f_1(x) + p_2 f_2(x)} \Rightarrow \delta_1(x) \geq \delta_2(x)$

3.2 Clasificaciones de 2 poblaciones multivariadas normales:

- Cada una es normal. $f_i(x) = \dfrac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)\right]$

→ Con $\Sigma_1 = \Sigma_2 = \Sigma$.

Si $(\mu_1-\mu_2)'\Sigma^{-1}x_0 - \frac{1}{2}(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1+\mu_2) \geq \ln\left(\dfrac{C(1|2)p_2}{C(2|1)p_1}\right)$

entonces $x_0 \in \pi_1$. Si no, entonces $x_0 \in \pi_2$

En general, si $(\bar{x}_1-\bar{x}_2)'S_{pool}^{-1}x_0 - \frac{1}{2}(\bar{x}_1-\bar{x}_2)'S_{pool}^{-1}(\bar{x}_1+\bar{x}_2) \geq \ln\left(\dfrac{C(1|2)\,p_2}{C(2|1)\,p_1}\right)$

entonces $x_0 \in \pi_1$. Si no, entonces $x_0 \in \pi_2$

$$S_{pool} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

→ Fisher's Approach: $\hat{y}_0 = (\bar{x}_1-\bar{x}_2)'S_{pool}^{-1}x_0$

$\hat{m} = \frac{1}{2}(\bar{x}_1-\bar{x}_2)'S_{pool}^{-1}(\bar{x}_1+\bar{x}_2)$

- Si $\hat{y}_0 \geq \hat{m} \Rightarrow x_0 \in \pi_1$ • Si $\hat{y}_0 < \hat{m} \Rightarrow x_0 \in \pi_2$.

→ Con $\Sigma_1 \neq \Sigma_2$.

Si $-\frac{1}{2}x_0'(\Sigma_1^{-1}-\Sigma_2^{-1})x_0 + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})x_0 - K \geq \ln\left(\dfrac{C(1|2)\,p_2}{C(2|1)\,p_1}\right)$

$\Rightarrow x_0 \in \pi_1$, Si no, $x_0 \in \pi_2$, $K = \frac{1}{2}\ln\left(\dfrac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\mu_1'\Sigma_1^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2)$

En general,

Si $-\frac{1}{2}x_0'(S_1^{-1}-S_2^{-1})x_0 + (\bar{x}_1'S_1^{-1} - \bar{x}_2'S_2^{-1})x_0 - K \geq \ln\left(\dfrac{C(1|2)}{C(2|1)}\dfrac{p_2}{p_1}\right)$

$\Rightarrow x_0 \in \pi_1$. Si no, $x_0 \in \pi_2$, $K = \frac{1}{2}\ln\left(\dfrac{|S_1|}{|S_2|}\right) + \frac{1}{2}(\bar{x}_1'S_1^{-1}\bar{x}_1 - \bar{x}_2'S_2^{-1}\bar{x}_2)$

3.3 Evaluar la clasificación:

real \ Pred	π_1	π_2
π_1	n_{1C}	n_{2M}
π_2	n_{2M}	n_{2C}

$APER = \dfrac{n_{1M} + n_{2M}}{n_1+n_2}$

n_{1C} → Elementos en π_1 que son de π_1

n_{2C} → Elementos en π_2 que son de π_2

$n_{1M} = n_1 - n_{1C}$ $\quad n_{2M} = n_2 - n_{2C}$

3.4. Varias poblaciones: para más de 2 categorías los costos siempre son =

$d_i(x)$. $\quad P(k|i) = \int_{R_k} f_i(x)\,dx$ $\quad ECM = \left(\sum P(k|1)C(k|1)\right)\sum p_i$

→ clasificación con ECM. $x_0 \in \pi_k$ si $p_k f_k(x) \geq p_i f_i(x) \quad \forall i \neq k$

QDA: $d_i(x) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) + \ln p_i$

QDA_sample: $d_i(x) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(x-\bar{x}_i)'S_i^{-1}(x-\bar{x}_i) + \ln p_i$

LDA: $d_i(x) = \mu_i'\Sigma^{-1}x - \frac{1}{2}\mu_i'\Sigma^{-1}\mu_i + \ln p_i$

LDA_sample: $d_i(x) = \bar{x}_i'S_{pool}^{-1}x - \frac{1}{2}\bar{x}_i'S_{pool}^{-1}\bar{x}_i + \ln p_i$

$S_{pool} = \dfrac{(n_1-1)S_1 + \dots + (n_p-1)S_p}{n_1+\dots+n_p-p}$ $\quad x \in \pi_k$ si $\max\{d_i(x)\} = d_k(x)$

4. Clustering: hay 2 metodos → K-means
→ Jerarquico.

4.1 K-means. En el taller preparcial 3 también dice el paso a paso

Algoritmo: 1. Asigne una etiqueta random de 1 a K a cada observación. Estos son clusters iniciales.
2. Itere hasta que dejen de cambiar los clusters:
a. Para cada K cluster saque un centroide, a partir de las medias segun la etiqueta.
b. Cambie los clusters al que este más cerca con distancia euclideana.

Ej K=3. Diapositiva 17

4.2 Jerarquico: → complete. max
→ Single. min
→ Average. Avg

Algoritmo: 1. Empiece con n observaciones y las distancias con la matriz de disimilitud.
2. For i=n, n-1, ..., 2:
a. Fusione las 2 obs mas cerca y saque las nuevas obs y sus distancias con las demas
b. saque la nueva matriz de disimilitud.

* la matriz de disimilitud se saca con distancia euclidiana, si son muchos datos se usa matriz de correlación R