Tarea 1 Reconocimiento de Patrones

Yamil Ernesto Morfa Avalos

4 de febrero de 2021

A. Preguntas Cortas

2. Para un estudio se mide la temperatura en diferentes partes del cuerpo de una muestra de personas. Un investigador expresa todas las temperaturas en grados Celsius. Otro investigador convierte primero todas estas temperaturas a grados Fahrenheit.

Supongamos que se miden k partes del cuerpo en n personas. Cada muestra de temperatura tomada en grados Celsius $x_i = [x_{i1}, x_{i2}, ..., x_{ik}] \in \mathbb{R}^k$ con $i = 1, 2, ..., n$ se representa como la fila de la matriz

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Dado que la relación entre grados Celsius y grados Fahrenheit es de la siguiente forma

$$^{\circ}C * \frac{9}{5} + 32 =^{\circ} F$$

tenemos que Z, que representa las temperaturas en grados Fahrenheit, se puede escribir de la forma

$$z_{ij} = \frac{9}{5} x_{ij} + 32 \equiv Z = \frac{9}{5} X + 32\mathbf{1}_{n\times k}$$

Luego

$$Cov\left(Z\right) = E\left[\left(Z - E\left[Z\right]\right)\left(Z - E\left[Z\right]\right)^T\right] = E\left[ZZ^T\right] - E\left[Z\right]\left(E\left[Z\right]\right)^T$$

$$= E\left[\left(\frac{9}{5}X + 32\mathbf{1}_{n\times k}\right)\left(\frac{9}{5}X + 32\mathbf{1}_{n\times k}\right)^T\right] - E\left[\frac{9}{5}X + 32\mathbf{1}_{n\times k}\right]\left(E\left[\frac{9}{5}X + 32\mathbf{1}_{n\times k}\right]\right)^T =$$

$$= E\left[\left(\frac{9}{5}\right)^2 XX^T + \left(32\right)^2 \mathbf{1}_{n\times k}\mathbf{1}_{n\times k}^T + 32\frac{9}{5}X\mathbf{1}_{n\times k}^T + 32\frac{9}{5}\mathbf{1}_{n\times k}X^T\right] -$$

$$\left(\frac{9}{5}E\left[X\right] + 32\mathbf{1}_{n\times k}\right)\left(\frac{9}{5}E\left[X\right] + 32\mathbf{1}_{n\times k}\right)^T =$$

$$= \left(\frac{9}{5}\right)^2 E\left[XX^T\right] - \left(\frac{9}{5}\right)^2 E\left[X\right]\left(E\left[X\right]\right)^T = \left(\frac{9}{5}\right)^2 \left(E\left[XX^T\right] - E\left[X\right]\left(E\left[X\right]\right)^T\right) = \left(\frac{9}{5}\right)^2 Cov\left(X\right)$$

Entonces la relación entre sus matrices de covarianza es:

$$Cov\left(Z\right) = \left(\frac{9}{5}\right)^2 Cov\left(X\right)$$

Supongamos ahora que ambos deciden proyectar en la misma dirección de máxima varianza. ¿obtendrán las mismas direcciones de proyección?

Primero comprobemos que hay una cierta relación entre los eigenvalores de $Cov\left(Z\right)$ y $Cov\left(X\right)$. Supongamos que λ es un eigenvalor de $Cov\left(Z\right)$, entonces

$$\det\left[Cov\left(Z\right) - \lambda I\right] = 0 \Longrightarrow \det\left[\left(\frac{9}{5}\right)^2 Cov\left(X\right) - \lambda I\right] = 0 \Longrightarrow \left(\frac{9}{5}\right)^{2n} \det\left[Cov\left(X\right) - \left(\frac{5}{9}\right)^2 \lambda I\right] = 0$$

$$\Longrightarrow \det\left[Cov\left(X\right) - \left(\frac{5}{9}\right)^2 \lambda I\right] = 0$$

Pero esto es equivalente a decir que $\gamma = \left(\frac{5}{9}\right)^2 \lambda$ es valor propio de $Cov\left(X\right)$. Luego, como se mantiene una relación de orden, si seleccionamos λ^* el mayor valor propio (máxima varianza) para $Cov\left(Z\right)$ tenemos que $\gamma^* = \left(\frac{5}{9}\right)^2 \lambda^*$ es la máxima varianza para $Cov\left(X\right)$. Ahora es fácil notar que la dirección de proyección para ambas es la misma:

$$Cov\left(Z\right) d = \lambda^* d \Longleftrightarrow \left(\frac{9}{5}\right)^2 Cov\left(X\right) d = \lambda^* d \Longleftrightarrow Cov\left(X\right) d = \left(\frac{5}{9}\right)^2 \lambda^* d \Longleftrightarrow Cov\left(X\right) d = \gamma^* d$$

3. Supongamos que $X = \left(X_1, X_2\right),\ Var\left(X_1\right) = Var\left(X_2\right) = 1.$

$$Cov\left(X\right) = \begin{bmatrix} 1 & E\left[\left(X_1 - E\left[X_1\right]\right)\left(X_2 - E\left[X_2\right]\right)\right] \\ E\left[\left(X_2 - E\left[X_2\right]\right)\left(X_1 - E\left[X_1\right]\right)\right] & 1 \end{bmatrix}$$

Sea $k = E\left[\left(X_1 - E\left[X_1\right]\right)\left(X_2 - E\left[X_2\right]\right)\right] = E\left[\left(X_2 - E\left[X_2\right]\right)\left(X_1 - E\left[X_1\right]\right)\right]$, tenemos:

$$\det\left\{\begin{bmatrix} 1-\lambda & k \\ k & 1-\lambda \end{bmatrix}\right\} = 0 \Longrightarrow \left(1-\lambda\right)^2 - k^2 = 0 \Longrightarrow \left(1-\lambda+k\right)\left(1-\lambda-k\right) = 0$$

Luego

$$\lambda_{1,2} = 1 \pm k$$

Sea $\lambda = \lambda_1 = 1 + k$ tenemos

$$\begin{bmatrix} -k & k \\ k & -k \end{bmatrix} v_1 = 0_2 \sim \begin{bmatrix} -k & k \\ 0 & 0 \end{bmatrix} v_1 = 0_2 \Longrightarrow v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Sea $\lambda = \lambda_2 = 1 - k$ tenemos

$$\begin{bmatrix} k & k \\ k & k \end{bmatrix} v_2 = 0_2 \sim \begin{bmatrix} k & k \\ 0 & 0 \end{bmatrix} v_2 = 0_2 \Longrightarrow v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Notemos qué si $k > 0$ entonces λ_1 sería la máxima varianza y la dirección de proyección estría dad por $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

5. Sabemos qué PCA es equivalente a resolver el problema:

$$\max_{l} \frac{l^T Cov\,(X)\,l}{l^T l}$$

usando la descomposición en valores singulares de la matriz simétrica $Cov\,(X) = U\Lambda U^T$ tenemos qué:

$$\max_{\|l\|\neq 0} \frac{l^T U \Lambda U^T l}{l^T U U^T l} = \max_{\|l\|\neq 0} \frac{l^T U \Lambda^{1/2} U^T U \Lambda^{1/2} U^T l}{l^T U U^T l}$$

Sea $y^T = l^T U$, tenemos que:

$$\max_{\|y\|\neq 0} \frac{y^T \Lambda^{1/2} U^T U \Lambda^{1/2} y}{y^T y} = \max_{\|y\|\neq 0} \frac{y^T \Lambda y}{y^T y} = *$$

Si agreguemos además la restricción de que y tenga que ser ortogonal al primer vector propio $u_{:,1}$. Dado que $\{u_{:,1}, u_{:,2}, ..., u_{:,n}\}$ es una base ortogonal, entonces $y = \sum_{i=1}^n \alpha_i u_{:,i}$ y como $y^T u_{:,1} = 0 \Longrightarrow \alpha_1 = 0$

$$* = \max_{\|y\|\neq 0} \frac{\left(\sum_{i=2}^n \alpha_i u_{:,i}\right)^T \Lambda \left(\sum_{i=2}^n \alpha_i u_{:,i}\right)}{\left(\sum_{i=2}^n \alpha_i u_{:,i}\right)^T \left(\sum_{i=2}^n \alpha_i u_{:,i}\right)} = \max_{\|y\|\neq 0} \frac{\sum_{i=2}^n \lambda_i \alpha_i^2 u_{:,i}^T u_{:i}}{\sum_{i=2}^n \alpha_i^2 u_{:,i}^T u_{:i}} \geq \max_{\|y\|\neq 0} \frac{\lambda_2 \sum_{i=2}^n \alpha_i^2 u_{:,i}^T u_{:i}}{\sum_{i=2}^n \alpha_i^2 u_{:,i}^T u_{:i}} = \lambda_2$$

entonces para $y = [0, 1, 0, ..., 0]$ tenemos que se alcanza el máximo para este problema. Ahora regresando al valor original tenemos que:

$$y^T = l^T U \Longrightarrow Cov(X) l^T = y^T U^T \equiv l = Uy = u_{:,2}$$

que es el segundo vector propio de $Cov(X)$.

B. Análisis de datos

1. En deport.dat se encuentran, por país, el mejor tiempo obtenido (hasta 1984) en diferentes pruebas de pista (ordenadas según distancia). En la figura 1 y 2 se muestra una grafica de las matrices de dispersión de los datos y la correlación entre estos respectivamente.

Figura 1:

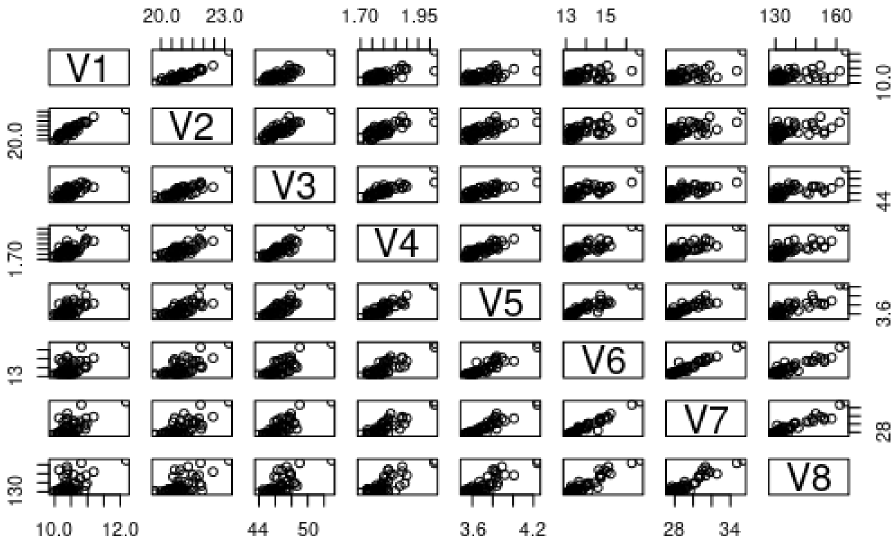
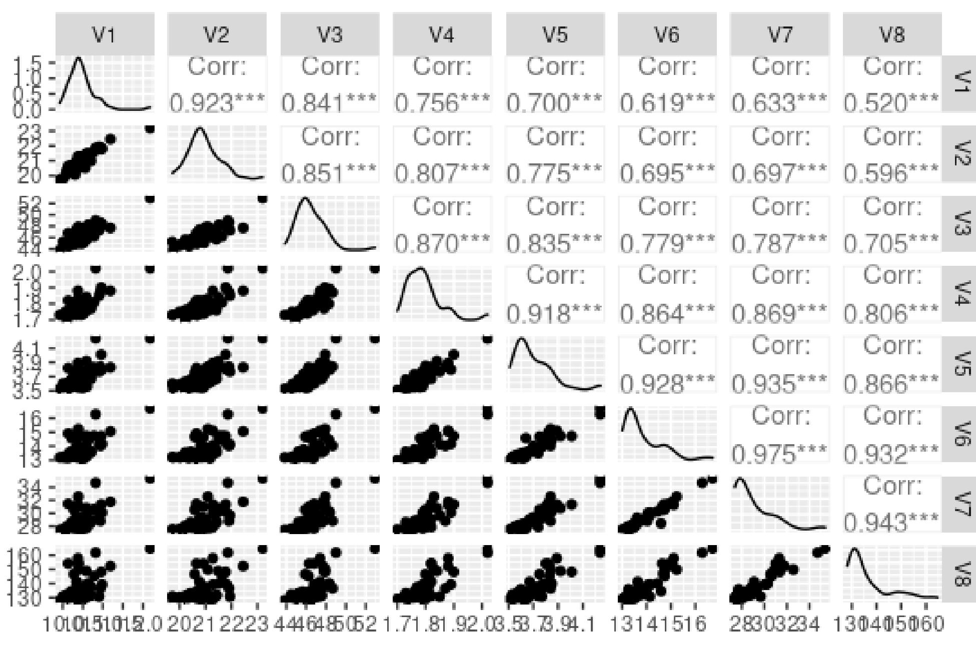
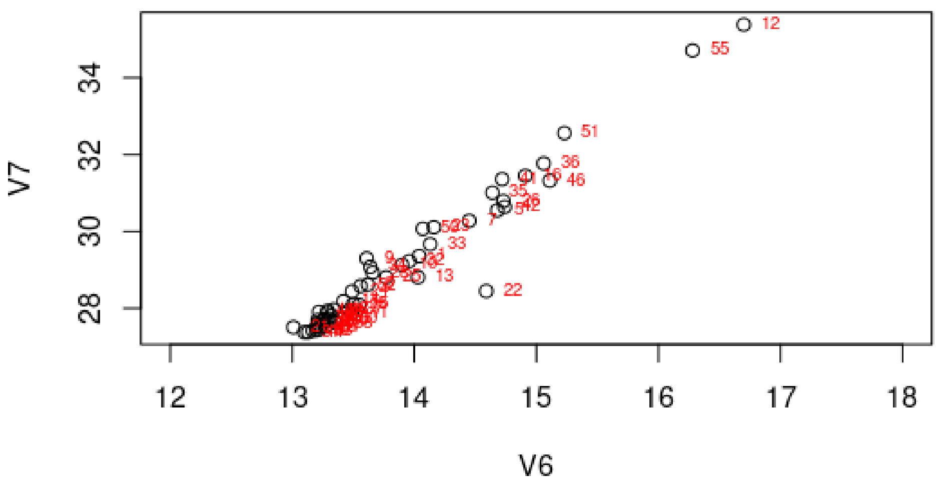


Figura 2:



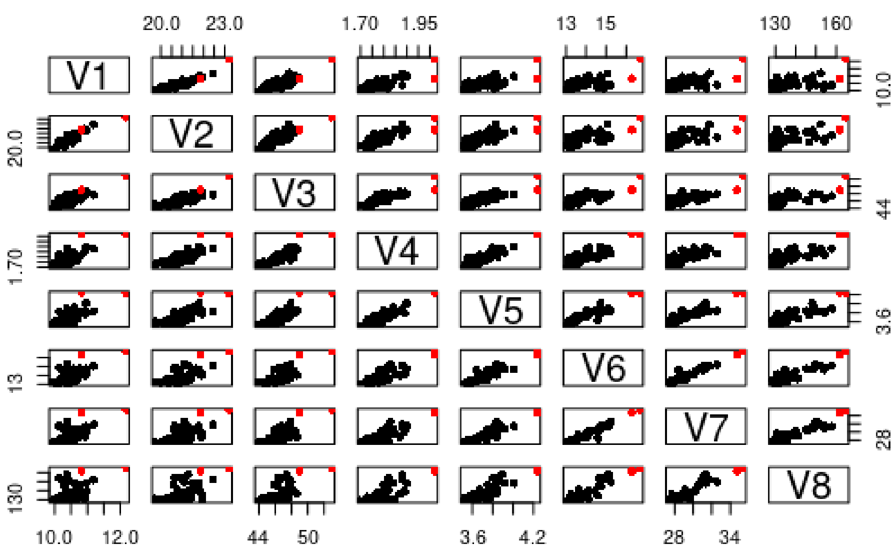
El valor mas alto de correlación entre los datos, con un valor de 0.975 se encuentra entre las variables representadas por V6 y V7. Veamos más de cerca este caso en la Figura 3

Figura 3:



Podemos notar que hay datos atípicos, en los indices 12 y 55 que representan a cookis y wsamoa respectivamente, veamos si esto sucede en todas las variables

Figura 4:



En la figura 5 se muestra la salida del PCA para los datos. Podemos notar que las variables que influyen en la primera componente son V3, V7, V8 siendo V8 (que representa la carrera de mayor duración) la mas influyente de estas.

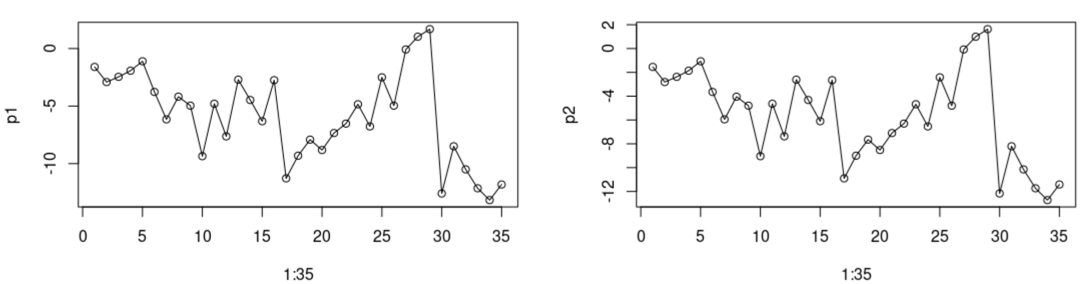
Figura 5:

```
Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
V1         0.211         0.359  0.190  0.887
V2         0.359         0.834        -0.410
V3  0.111  0.828  0.378 -0.396
V4                                            0.261 -0.965
V5                                            0.959  0.262
V6         0.130 -0.336        -0.909  0.184
V7  0.181  0.299 -0.849 -0.135  0.364
V8  0.973 -0.181  0.142

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.125  0.125  0.125  0.125  0.125  0.125  0.125  0.125
Cumulative Var  0.125  0.250  0.375  0.500  0.625  0.750  0.875  1.000
>
```

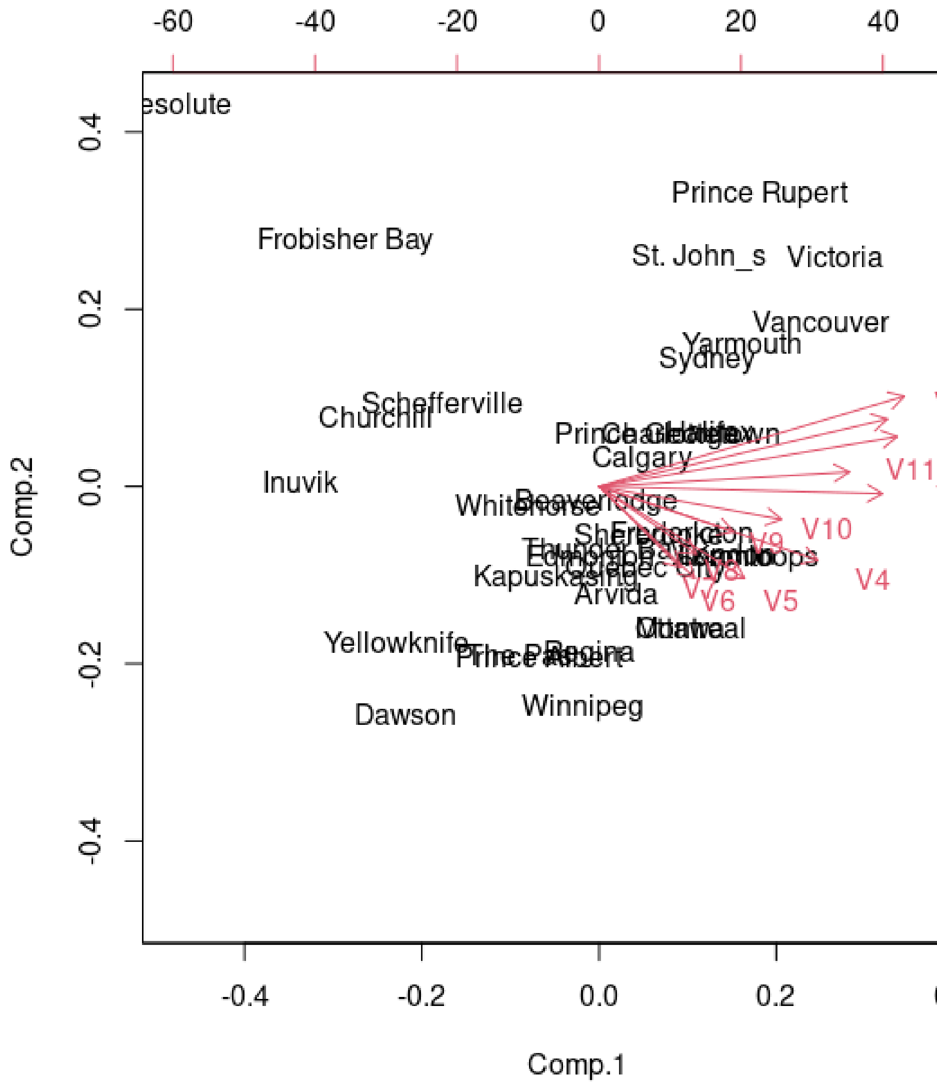
2. Considera los datos oef2.data. Se trata de los promedios mensuales de la temperatura (en Celsius) en 35 estaciones canadienses de monitoreo. El interés es comparar las estaciones entre sí en base de sus curvas de temperatura. Considerando las 12 mediciones por estación como un vector X. Apliquemos PCA, en la figura siguiente se muestra las proyecciones de los datos sobre la primera y segunda componente respectivamente

Figura 6:



Ahora si proyectamos los datos al plano formado por las primeras dos componentes principales y agregamos a cada punto su etiqueta de estación como se muestra en la figura 7

Figura 7:


-60
-40
-20
0
20
40
esolute
Prince Rupert
Frobisher Bay
St. John_s
Victoria
Vancouver
Yarmouth
Sydney
Schefferville
Churchill
Calgary
Inuvik
Whitehorse
Beaverlodge
V11
V10
V9
V4
V5
V6
Kapuskasing
Arvida
Yellowknife
Regina
Prince Albert
Winnipeg
Dawson
Comp.2
Comp.1
0.4
0.2
0.0
-0.2
-0.4