

Tarea 1 - Reconocimiento estadístico de patrones

Erick Salvador Alvarez Valencia

CIMAT A.C.,
`erick.alvarez@cimat.mx`

Resumen En el presente reporte se hablará sobre la resolución de los ejercicios de la primer tarea de reconocimiento de patrones, de la misma manera se presentará el código de los mismos y los resultados que fueron obtenidos.

1. Problema 1

En base del artículo de Jordan and Mitchell sobre Machine Learning: ¿Cómo definirías Machine Learning en 2-3 párrafos?

Machine learning es una rama derivada de la computación y la estadística que en los últimos años se ha popularizado bastante gracias a los avances que ha mostrado. Los dos principales objetivos de esta área son: Cómo hacer sistemas robustos que puedan aprender a hacer ciertas cosas tal como lo hace el hombre (leyes que gobiernan el aprendizaje en los humanos) y cómo hacer algoritmos que puedan mejorar a través de la experiencia (sean más precisos en cada ocasión), para esto último hay muchas técnicas basadas en optimización que buscan hacer que una función objetivo alcance un máximo o un mínimo en base a los errores cometidos durante la ejecución.

Machine learning también se puede ver como una serie de algoritmos que pueden identificar ciertos patrones desde una gran cantidad de datos en crudo usando técnicas estadísticas tales como clasificadores binarios Bayesianos, esto anterior con el fin de lograr sistemas de predicción que pueden ser muy útiles en áreas como: economía, medicina, comercio, climatología, entre otras muchas.

2. Problema 2

Para un estudio se mide la temperatura en diferentes posiciones del cuerpo de una muestra de personas. Un investigador expresa todas las temperaturas en grados Celcius. Otro investigador convierte primero todas estas temperaturas a grados Fahrenheit. Si ambos deciden hacer PCA, ¿obtendrán las mismas direcciones de proyección? Explica tu respuesta de manera formal (con una demostración).

Sabemos que la fórmula para convertir grados Celcius a Fahrenheit es: $^{\circ}F = 1,8^{\circ}C + 32$ la cual se puede ver como una combinación lineal de la variable C .

Suponiendo que el investigador A tiene sus datos en grados Celcius, sea X la matriz de dichos datos, por su parte, el investigador B convierte a los datos de la matriz X a una versión en Fahrenheit usando la fórmula anterior, esto lo podemos ver como: $Y = aX + b$ donde Y es la matriz resultante y a es un escalar y b una matriz. Podemos ver que le pasa a la matriz Y al aplicar PCA.

La matriz de covarianza de Y la definimos como $\Sigma = Cov(Y)$, sustituyendo por la definición de Y tenemos:

$$\Sigma = Cov(aX + b) \quad (1)$$

Podemos usar la siguiente propiedad de la covarianza:

$$Cov(CX + D) = CCov(X)C^T \quad (2)$$

Donde C es una matriz. Aplicando la propiedad anterior en nuestro caso nos queda:

$$\begin{aligned} Cov(Y) &= aCov(X)a^T \\ Cov(Y) &= a^2Cov(X) \end{aligned} \quad (3)$$

Como a es un escalar al final queda elevado al cuadrado.

Sea u_1 el vector propio asociado al valor propio más grande λ_1 con $\|u_1\|_2 = 1$, por lo anterior se puede ver que u_1 es un vector propio de la matriz X . Ahora, haciendo la proyección de dicho vector hacia la matriz de covarianza de Y tenemos:

$$\begin{aligned} u_1 \Sigma u_1^T &= \lambda_1 \\ u_1 Cov(Y) u_1^T &= \lambda_1 \\ u_1 a^2 Cov(X) u_1^T &= \lambda_1 \\ u_1 Cov(X) u_1^T &= a^{-2} \lambda_1 \end{aligned} \quad (4)$$

Podemos ver que finalmente el valor propio quedará multiplicando al recíproco de a^2 . Esto anterior pasaría de la misma forma con cualquier vector propio que hubiéramos tomado.

Como conclusión, aplicar PCA a la transformación de los datos de Celcius a Fahrenheit provocará un reescalamiento a los valores propios de la matriz de covarianza, pero como a^2 es positivo y los valores propios también lo son porque la matriz de covarianza es simétrica y definida positiva podemos ver que el orden de los autovalores no cambia y como los autovectores son los de la matriz X , estaremos haciendo las mismas proyecciones que se harían con la matriz X .

3. Problema 3

Sea X una v.a. multidimensional con matriz de covarianza $Cov(X)$ y $EX = 0$. Si l es un vector propio de $Cov(X)$ con valor propio λ y $Y = \langle l, X \rangle$, muestra que:

$$Cov(Y, X_i) = \lambda l_i$$

, con entrada i del vector l .

Como l es un vector propio de la matriz de covarianza tenemos que:

$$Cov(X)l = \lambda l \quad (5)$$

Ahora si hacemos lo mismo pero sólo tomando el i -ésimo vector de la matriz de covarianza, nos queda:

$$\begin{aligned} Cov(X)_i l &= \lambda l_i \\ E((X - E(X))(X - E(X))^T)_i l &= \lambda l_i \\ E(XX^T)_i l &= \lambda l_i \\ \sum_{j=1}^n E(X_i X_j l_j) &= \lambda l_i \end{aligned} \quad (6)$$

Por otra parte tenemos que:

$$\begin{aligned} Cov(Y, X_i) &= Cov(l^T X, X_i) \\ Cov(Y, X_i) &= E(l^T X X_i) - E(l^T X)E(X_i) \\ Cov(Y, X_i) &= E(l^T X X_i) - l^T E(X)E(X_i) \\ Cov(Y, X_i) &= E(l^T X X_i) \end{aligned} \quad (7)$$

Lo anterior se hizo por la hipótesis de $E(X) = 0$, ahora podemos ver el producto punto como una suma:

$$\begin{aligned} Cov(Y, X_i) &= E\left(\sum_{j=1}^n X_j l_j X_i\right) \\ Cov(Y, X_i) &= \sum_{j=1}^n E(X_i X_j l_j) \end{aligned} \quad (8)$$

De esta forma se ha llegado a lo que se tiene en (6) por lo cual queda demostrado.

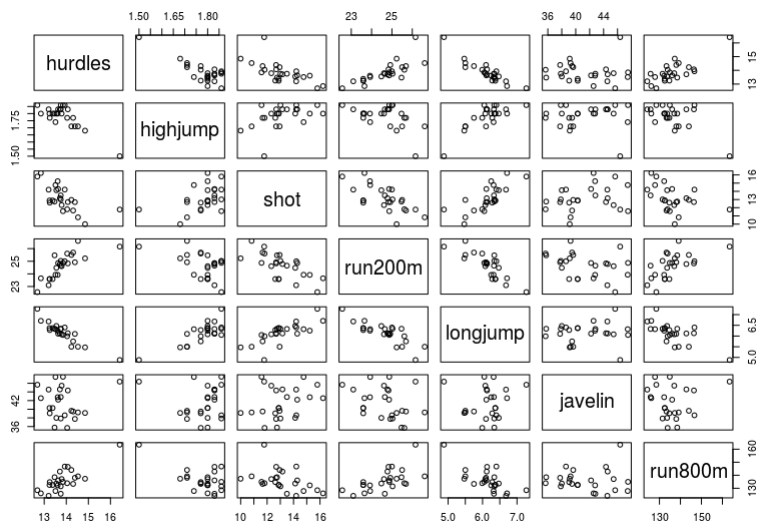
4. Problema 4

Un problema en un heptatlón es cómo convertir los resultados obtenidos en las diferentes pruebas en un puntaje final. En este ejercicio usamos PCA para este problema. En el archivo heptatlón se pueden consultar los tiempos y el puntaje final (score) de 25 atletas que participaron en el heptatlón durante los juegos olímpicos de Seoul. Visualiza los datos. Haz un análisis de componentes principales con los tiempos (sin score). ¿Hay una relación entre el score y las proyecciones sobre el primer CP?

Primero vamos a incluir la tabla de los datos y un plot de los mismos.

	hurdles	highjump	shot	run200m	longjump	javelin	run800m	score
Joyner-Kersey (USA)	12.69	1.86	15.80	22.56	7.27	45.66	128.51	7291
John (GDR)	12.85	1.80	16.23	23.65	6.71	42.56	126.12	6897
Behmer (GDR)	13.20	1.83	14.20	23.10	6.68	44.54	124.20	6858
Sablovskaitė (URS)	13.61	1.80	15.23	23.92	6.25	42.78	132.24	6540
Choubenkova (URS)	13.51	1.74	14.76	23.93	6.32	47.46	127.90	6540
Schulz (GDR)	13.75	1.83	13.50	24.65	6.33	42.82	125.79	6411
Fleming (AUS)	13.38	1.80	12.88	23.59	6.37	40.28	132.54	6351
Greiner (USA)	13.55	1.80	14.13	24.48	6.47	38.00	133.65	6297
Lajbnerova (CZE)	13.63	1.83	14.28	24.86	6.11	42.20	136.05	6252
Bouraga (URS)	13.25	1.77	12.62	23.59	6.28	39.06	134.74	6252
Wijnsma (HOL)	13.75	1.86	13.01	25.03	6.34	37.86	131.49	6205
Dimitrova (BUL)	13.24	1.80	12.88	23.59	6.37	40.28	132.54	6171
Scheider (SWI)	13.85	1.86	11.58	24.87	6.05	47.50	134.93	6137
Braun (FRG)	13.71	1.83	13.16	24.78	6.12	44.58	142.82	6109
Ruotsalainen (FIN)	13.79	1.80	12.32	24.61	6.08	45.44	137.06	6101
Yuping (CHN)	13.93	1.86	14.21	25.00	6.40	38.60	146.67	6087
Hagger (GB)	13.47	1.80	12.75	25.47	6.34	35.76	138.48	5975
Brown (USA)	14.07	1.83	12.69	24.83	6.13	44.34	146.43	5972
Mulliner (GB)	14.39	1.71	12.68	24.92	6.10	37.76	138.02	5746
Hautenauve (BEL)	14.04	1.77	11.81	25.61	5.99	35.68	133.90	5734
Kytola (FIN)	14.31	1.77	11.66	25.69	5.75	39.48	133.35	5686
Geremias (BRA)	14.23	1.71	12.95	25.50	5.50	39.64	144.02	5508
Hui-Ing (TAI)	14.85	1.68	10.00	25.23	5.47	39.14	137.30	5290
Jeong-Mi (KOR)	14.53	1.71	10.83	26.61	5.50	39.26	139.17	5289
Launa (PNG)	16.42	1.50	11.78	26.16	4.88	46.38	163.43	4566

(a) Figura 1. Tabla generada por los datos de Heptatlón.



(b) Figura 2. Plot de los datos de Heptatlón.

Posteriormente se aplicó PCA a los datos usando la función *textitprincomp* de la siguiente manera: $p \text{ j- princomp}(d[, -8])$ esto se hace de esa manera para excluir la columna **Score**. A continuación se muestran los resultados obtenidos:

```

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
hurdles              0.222  0.327 -0.807  0.425
highjump              0.140              0.985
shot                 -0.136 -0.884  0.425 -0.104
run200m              0.101  0.310  0.816  0.462
longjump             -0.185 -0.204  0.319  0.895 -0.142
javelin              -0.985  0.160
run800m              0.991      -0.117

      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
SS loadings    1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.143  0.143  0.143  0.143  0.143  0.143  0.143
Cumulative Var  0.143  0.286  0.429  0.571  0.714  0.857  1.000

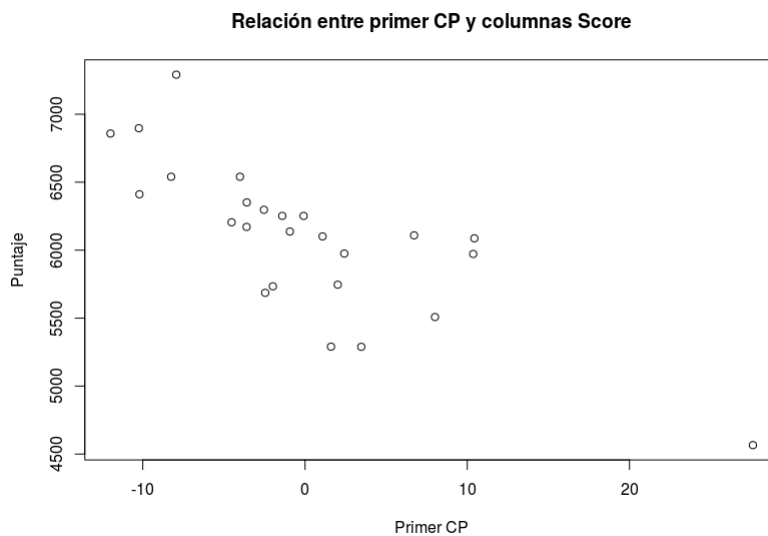
```

(c) Figura 3. Resultado del PCA aplicado a los datos.

```
> summary(p)
Importance of components:
              Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6   Comp.7
Standard deviation  8.1956429  3.5184227  1.3577008  0.573879298  0.317274394  0.1441496680  3.257782e-02
Proportion of Variance 0.8206996 0.1512566 0.0225230 0.004024012 0.001229952 0.0002538899 1.296768e-05
Cumulative Proportion 0.8206996 0.9719562 0.9944792 0.998503190 0.999733142 0.9999870323 1.000000e+00
```

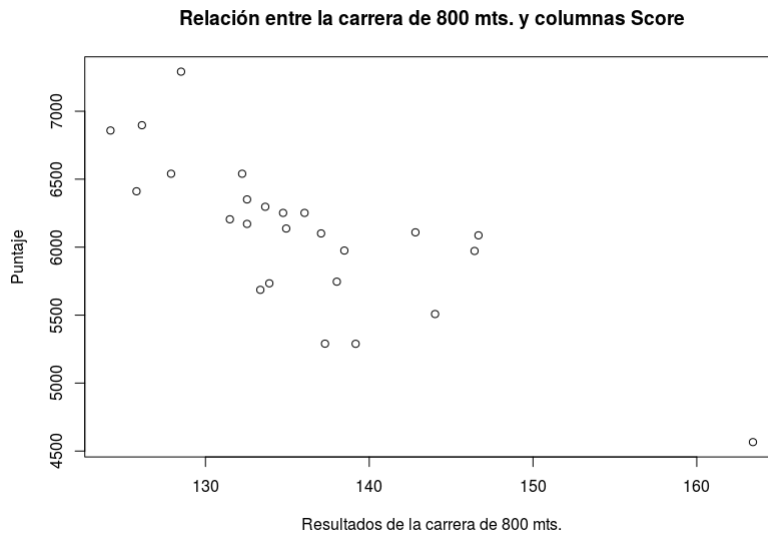
(d) Figura 4. Resultado del PCA aplicado a los datos.

En la Figura 4 podemos notar que el componente 1 es el que aporta el 82 % de la varianza de los datos y dicha varianza es aportada por la carrera de 800 mts., eso lo podemos ver en la Figura 3. Pero para apreciarlo mejor podemos ver la gráfica entre estos dos conjuntos de datos.



(e) Figura 5. Gráfica del primer componente contra el puntaje.

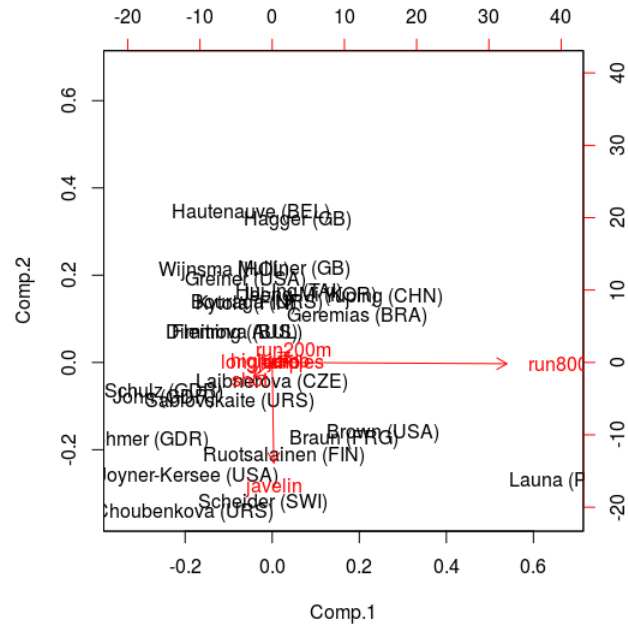
Podríamos decir que existe una relación lineal entre el primer componente y la columna *Score* de los datos de tal forma que entre menos tiempo hizo el atleta en ese evento más subió su puntaje final. Osea que ese es el evento que más peso tiene de todos. Aunque esto puede no quedar tan obvio, aún así podemos ver la misma relación graficando la columna de datos de 800 mts. con la misma columna de puntaje.



(f) Figura 6. Gráfica de los 800 mts. contra el puntaje.

De la Figura 6. notamos la misma relación lineal anterior vista en la gráfica de la Figura 5 y es por lo dicho que la carrera de 800 mts. es la que más variabilidad aporta.

Por último veremos el biplot generado por el PCA.



(g) Figura 7. Biplot generado por el PCA.

El biplot nos muestra las dos componentes principales que aportan más variabilidad, el cual es la carrera de 800 mts. y el lanzamiento de javalina, aunque cuesta ver la información con estos dos vectores. Puede quedarnos la duda del por qué el evento de 800 mts. es el que aporta la mayor variabilidad por lo cual otro enfoque que podemos hacer es trabajar con los datos normalizados usando la matriz de correlación en lugar de la de covarianza y hacer un PCA. A continuación se muestra el mismo análisis con la matriz de correlación:


```

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
hurdles  0.453 -0.158          0.783  0.380
highjump -0.377  0.248 -0.368  0.680          0.434
shot     -0.363 -0.289  0.676  0.124 -0.512          0.218
run200m  0.408  0.260          0.361 -0.650          -0.453
longjump -0.456          0.139  0.111  0.184  0.590 -0.612
javelin   -0.842 -0.472  0.121 -0.135          -0.173
run800m   0.375 -0.224  0.396  0.603  0.504 -0.156

      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
SS loadings  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var 0.143  0.143  0.143  0.143  0.143  0.143  0.143
Cumulative Var 0.143  0.286  0.429  0.571  0.714  0.857  1.000

```

(h) Figura 8. Resultados del PCA encontrados con la matriz de correlación.

```

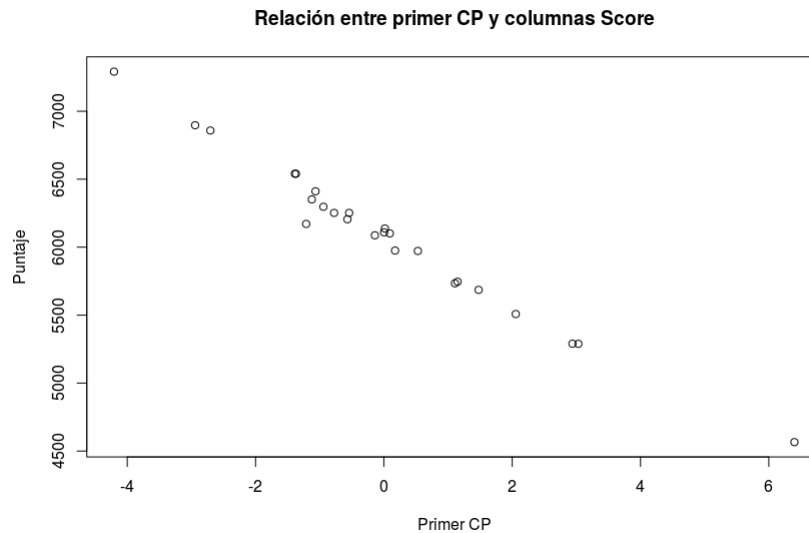
> summary(p)
Importance of components:
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6      Comp.7
Standard deviation  2.1119364  1.0928497  0.72181309  0.67614113  0.49524412  0.27010291  0.221361710
Proportion of Variance 0.6371822  0.1706172  0.07443059  0.06530955  0.03503811  0.01042223  0.007000144
Cumulative Proportion 0.6371822  0.8077994  0.88222998  0.94753952  0.98257763  0.99299986  1.000000000

```

(i) Figura 9. Resultados del PCA encontrados con la matriz de correlación.

En las Figuras 8 y 9 vemos los resultados encontrados usando PCA a la matriz de correlación de los datos originales. Lo primero que podemos notar aquí es que ahora la variabilidad no se ve tan sesgada al primer componente principal, de igual manera por cada CP la mayor parte de los eventos aportan más variabilidad en comparación a los resultados anteriormente obtenidos.

Ahora podemos graficar el primer CP en contra de la columna *Score* de los datos:



(j) Figura 10. Gráfica del primer CP (Correlación) contral el puntaje.

En la Figura 10 podemos ver más claramente la relación lineal existente entre el primer CP y los puntajes finales obtenidos con los atletas. Finalmente podemos decir que se obtuvo una mejor perspectiva de los datos al usar la matriz de correlación en lugar de la de covarianza la cual provocaba que hubiera un sesgo con la carrera de 800 mts. cuando la realidad era que todos los eventos aportaban cierta variabilidad al puntaje final.

A continuación se muestra el código empleado para el ejercicio:

```
library("GGally")
setwd("Indicar la carpeta donde se encuentra el archivo")
d <- read.table("hepatlon") donde se encuentra el archivo
print(d)
ggpairs(d[, -8], diag = list(continuous = "density"), axisLabels = "none")
p <- princomp(d[, -8], cor = T)
loadings(p)
summary(p)
biplot(p)
plot(d$run800m, d$score, xlab = "Resultados de la carrera de 800 mts.", ylab = "Puntaje")
plot(p$scores[, 1], d$score, xlab = "Primer CP", ylab = "Puntaje", title("Relación entre primer CP y columnas Score"))
```

5. Problema 5

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 , aplica un análisis de componentes principales. Como X representa (un muestreo de) una curva, este tipo de datos se llama datos funcionales. Interpreta y dibuja (como curva) los primeros dos componentes, p_1 , p_2 es decir grafica (i, p_{1i}) y (i, p_{2i}) . Agrupa e interpreta las estaciones en el biplot (ten en mente un mapa de Canada).

Primeramente aplicamos PCA a los datos y a continuación se mostrará el resumen obtenido:

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
[1,]	-0.410	0.396	0.185	-0.397	-0.319	-0.158		-0.132	0.420	0.173	0.354	
[2,]	-0.401	0.219	-0.446	-0.147	0.381	-0.124	0.465		-0.178	0.324	-0.217	
[3,]	-0.381		-0.336	0.101		0.620		-0.140		-0.431	0.357	
[4,]	-0.294	-0.324	-0.372	0.268			-0.515		0.434	0.284	-0.247	
[5,]	-0.195	-0.403	-0.212		-0.373	-0.457		0.267	-0.441		0.350	-0.127
[6,]	-0.128	-0.401		-0.381	-0.345		0.293	-0.119	0.162	-0.328	-0.449	0.347
[7,]	-0.109	-0.355	0.241	-0.338	0.194	0.291	0.123	0.231	0.103	0.197		-0.665
[8,]	-0.131	-0.296	0.278	-0.205	0.440	0.135	-0.141	0.184		0.153	0.354	0.598
[9,]	-0.181	-0.198	0.242		0.147	-0.145	-0.240	-0.802	-0.315			-0.151
[10,]	-0.245	-0.145	0.230	0.443	0.306	-0.393	0.344		0.382	-0.353	0.150	
[11,]	-0.336		0.441	0.476	-0.352	0.272	0.200	0.103	-0.213	0.348	-0.179	0.122
[12,]	-0.387	0.296	0.165	-0.116	0.148	-0.123	-0.417	0.361	-0.268	-0.425	-0.352	

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
Cumulative Var	0.083	0.167	0.250	0.333	0.417	0.500	0.583	0.667	0.750	0.833	0.917	1.000

(k) Figura 11. Resultados del PCA aplicado a los datos.

```
> summary(p)
```

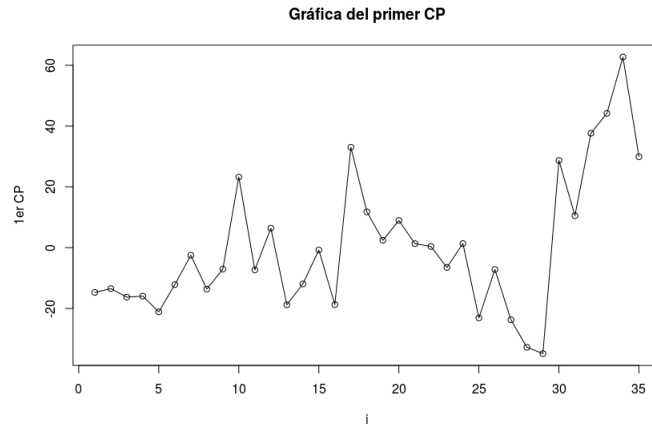
Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation	22.1502625	6.76867175	2.9974156	1.585358236	0.7058510606	0.6640082517	0.3837745104	0.3198832357	0.2346783515
Proportion of Variance	0.8932766	0.08341326	0.0163577	0.004575968	0.0009070983	0.0008027405	0.0002681516	0.0001862993	0.0001002707
Cumulative Proportion	0.8932766	0.97668985	0.9930475	0.997623515	0.9985306135	0.9993333539	0.9996015056	0.9997878048	0.9998880756

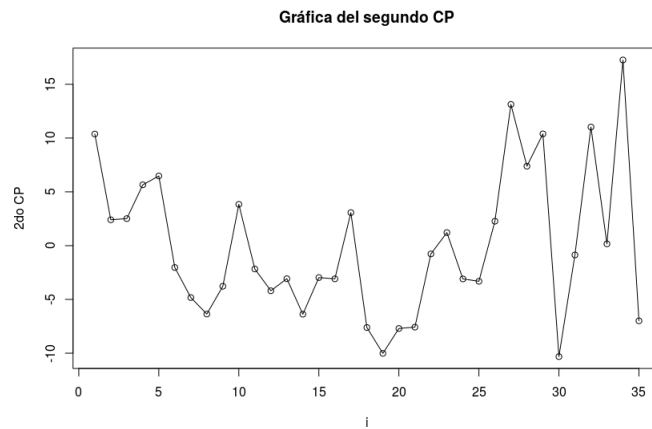
	Comp.10	Comp.11	Comp.12
Standard deviation	1.678721e-01	0.1503175110	1.034328e-01
Proportion of Variance	5.130801e-05	0.0000411384	1.947802e-05
Cumulative Proportion	9.999394e-01	0.9999805220	1.000000e+00

(l) Figura 12. Resultados del PCA aplicado a los datos.

En la Figura 12 podemos ver que más del 90% de la variabilidad de los datos es obtenida con los dos primeros CP y en la Figura 11 vemos que todas las estaciones aportan cierta variabilidad en general, pero en el primer CP podemos ver que las dos primeras mediciones son las de mayor variabilidad, en el caso del segundo CP pasa con las mediciones 4 y 5. Ahora se mostrará una grafica de los dos primeros componentes principales obtenidos:



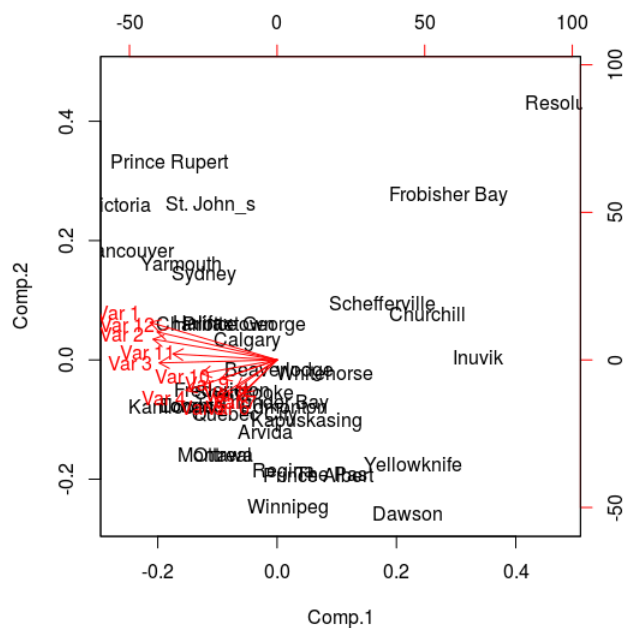
(m) Figura 13. Curva obtenida del primer CP.



(n) Figura 14. Curva obtenida del segundo CP.

En las Figuras 13 y 14 podemos observar las curvas obtenidas por los dos primeros componentes principales. Lo que se puede notar es que en la Figura 13 la cual representa al primer CP, las estaciones que están por arriba del 20 son las que en promedio tienen las temperaturas más extremas en las dos primeras mediciones. En cambio para la Figura 14 la cual representa la curva del segundo CP podríamos ver que pese a que existen muchos picos y valles en la gráfica no podemos concluir que esta muestra la misma información que la anterior pese a que la variabilidad aportada por este componente no es tanta como el primero, por lo que si queremos ver un gran cambio en las temperaturas de las estaciones

dictado por esta gráfica no vamos a tenerlo muy claro.
 Finalmente se mostrará el biplot generado por el PCA para mostrar las últimas conclusiones:



(ñ) Figura 15. Biplot del PCA.



(o) Figura 16. Mapa de Canadá.

El análisis que se hizo del biplot fue que en promedio, las estaciones fueron colocadas acorde a sus temperaturas medias. Posteriormente se hizo uso de un mapa de Canadá el cual se muestra en la Figura 16 para ver cuál era la relación de las estaciones con la ubicación obtenida de las mismas por el biplot. Finalmente se vió que en promedio, las estaciones que pertenecen a un mismo estado de Canada estan ubicadas muy cerca ya que se puede pensar que sus temperaturas no varían tanto.

A continuación se añade el código usado para tratar el problema:

```
setwd("Indicar_la_carpeta_donde_se_encuentra_el_archivo")
temp <- matrix(scan("oef2.data.dat"), 35, 12, byrow = T)

nombrestaciones <- c("St._John_s", "Charlottetown", "Halifax",
                     "Sydney", "Yarmouth", "Fredericton",
                     "Arvida", "Montreal", "Quebec_City",
                     "Schefferville", "Sherbrooke", "Kapuskasing",
                     "London", "Ottawa", "Thunder_Bay",
                     "Toronto", "Churchill", "The_Pas",
                     "Winnipeg", "Prince_Albert", "Regina",
```

```

"Beaverlodge", "Calgary", "Edmonton",
"Kamloops", "Prince_George", "Prince_Rupert",
"Vancouver", "Victoria", "Dawson",
"Whitehorse", "Frobisher_Bay", "Inuvik",
"Resolute", "Yellowknife")

rownames(temp) <- nombresestaciones
p <- princomp(temp)
print(temp)
loadings(p)
summary(p)
biplot(p)
i <- seq(1, length(p$scores[,1]))
print(p$scores)
j <- seq(1, length(p$scores[,2]))
plot(i, p$scores[,1], xlab = "i", ylab = "1er_CP", title(" Grafica_del_primer_CP"))
lines(i, p$scores[,1])
plot(j, p$scores[,2], xlab = "i", ylab = "2do_CP", title(" Grafica_del_segundo_CP"))
lines(j, p$scores[,2])

```

6. Problema 6

Sea K una matriz simétrica y positiva definida con rango r . Un resultado clásico de Rao dice que la matriz simétrica B de rango $p < r$ que minimiza

$$\|K - B\|_F$$

está dado por

$$B = \sum_{i=1}^p \lambda_i v_i v_i^T$$

, donde las v_i corresponden a los vectores propios ordenados de mayor a menor segun valor propio λ_i de K y $\|\cdot\|_F$ es la norma de Frobenius.

Verifica que en el mínimo

$$\|K - B\|_F^2 = \sum_{i=p+1}^r \lambda_i^2$$

Para algún vector propio v_i de la matriz K podemos ver qué pasa si multiplicamos $(K - B)^T(K - B)v_i$:

$$(K - B)^T(K - B)v_i = K^T K v_i - K^T B v_i - B^T K v_i + B^T B v_i \quad (9)$$

Como K y B son matrices simétricas tenemos que $K = K^T$ y $B = B^T$.

$$(K - B)^T(K - B)v_i = KKv_i - KBv_i - BKv_i + BBv_i \quad (10)$$

Lo primero que podemos notar de lo anterior es que como v_i es un vector propio de K se cumple: $Kv_i = \lambda v_i$. Ahora tenemos que ver que pasa con Bv_i y para ello usamos la definición de B :

$$Bv_i = \sum_{j=1}^p \lambda_j v_j v_j^T v_i \quad (11)$$

Tenemos que por la ortogonalidad de los vectores v , el producto $v_j^T v_i$ será cero si $j \neq i$, caso contrario será uno si $j = i$. Ahora, como estamos trabajando con los vectores propios de la matriz K se sabe que podemos escoger algún vector v_i donde $i > p$ cuando tengamos ese caso el producto $Bv_i = 0$ ya que para algún $j \in [1, p]$ no se podrá alcanzar a i .

Con lo anterior nos queda:

1. Si $i \leq p$:

$$(K - B)^T(K - B)v_i = \lambda_i^2 v_i - \lambda_i^2 v_i - \lambda_i^2 v_i + \lambda_i^2 v_i = 0$$

2. Si $i > p$:

$$(K - B)^T(K - B)v_i = \lambda_i^2 v_i$$

Ahora, sabemos que para algún vector propio v_i de K los autovalores de la matriz $(K - B)^T(K - B)$ son: 0 con multiplicidad p si $i \leq p$, λ_i^2 si $i > p$. Ahora se sabe que $\sigma_i = \sqrt{\alpha_i}$ es un valor singular de la matriz $(K - B)$ donde $\alpha_i = \lambda_i^2$ por lo que nos queda $\sigma_i = \lambda_i$ si $i > p$.

Para concluir, sabemos que la norma Frobenius se puede definir como:

$$\|A\|_F = \sqrt{\sum_{i=1}^{\min\{n,m\}} \sigma_i^2}$$

con $A \in R^{n \times m}$. Aplicando esta definición con la matriz $K - B$ nos queda

$$\|(K - B)\|_F^2 = \sum_{i=p+1}^r \lambda_i^2 \quad (12)$$