ANALÍTICA AVANZADA DE DATOS: ANÁLISIS DISCRIMINANTE LINEAL

A. Alejandra Sánchez Manilla
asanchezm.q@gmail.com

Reducción de dimensionalidad

En Aprendizaje Automático y Estadística, la Reducción de Dimensionalidad es el proceso de reducir el número de variables aleatorias consideradas mediante la obtención de un conjunto de variables principales. Puede dividirse en selección y extracción de características.

Dos algoritmos principales en la reducción de la dimensionalidad son:

- *Análisis de componentes principales (CPA)*
- *Análisis discriminante lineal (LDA)*

Análisis de Componentes Principales(PCA)

Principal Component Analysis (PCA)

Cuando trabajamos con muchas variables, podemos presentarnos con algunos problemas cómo:

- *Comprender la relación entre cada variable*
- *Tenemos tantas variables que corremos el riesgo de sobrescribir el modelo o violar alguna táctica de modelización*

Entonces, **¿Cómo tomamos todas las variables que hemos recopilado y centrarnos sólo en algunas de ellas?**

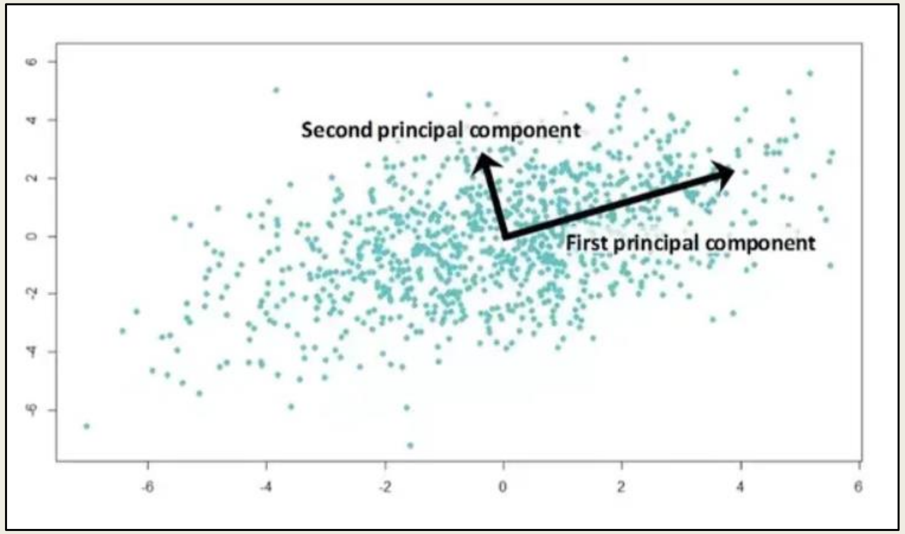
Análisis de Componentes Principales(PCA)

En términos técnicos, se trata de reducir la dimensión del espacio de las características"

Al reducir la dimensión del espacio de las características, tiene menos relaciones entre variables que considerar y menos probabilidad de sobrecargar el modelo, a esto también se le llama "reducción de dimensionalidad"

Hay muchas formas de reducir la dimensionalidad, pero la mayoría se dividen en dos categorías:

- *Eliminación de características*
- *Extracción de características*



Análisis de Componentes Principales(PCA)

- **Eliminación de características:** reducimos el espacio de características eliminándolas. Las ventajas del método de eliminación de características incluyen características de simplicidad y mantenibilidad. También eliminamos los beneficios que aportarían las variables eliminadas

- **Extracción de características:** PCA es una técnica de extracción de características. Así que combina nuestras variables de entrada de una manera específica, entonces podemos eliminar las variables "menos importantes" sin dejar de conservar las partes más valiosas de todas las variables

Análisis de Componentes Principales(PCA)

¿Cuándo debemos utilizar la PCA?

1. *¿Desea reducir el número de variables, pero no puede identificar las variables que desea eliminar por completo?*
2. *¿Desea asegurarse de que sus variables son independientes entre sí?*
3. *¿Necesitamos que la variable independiente sea menos interpretable?*

Análisis de Componentes Principales(PCA)

¿Cómo funciona el análisis de componentes principales?

Vamos a calcular una matriz que resuma cómo se relacionan entre sí todas nuestras variables

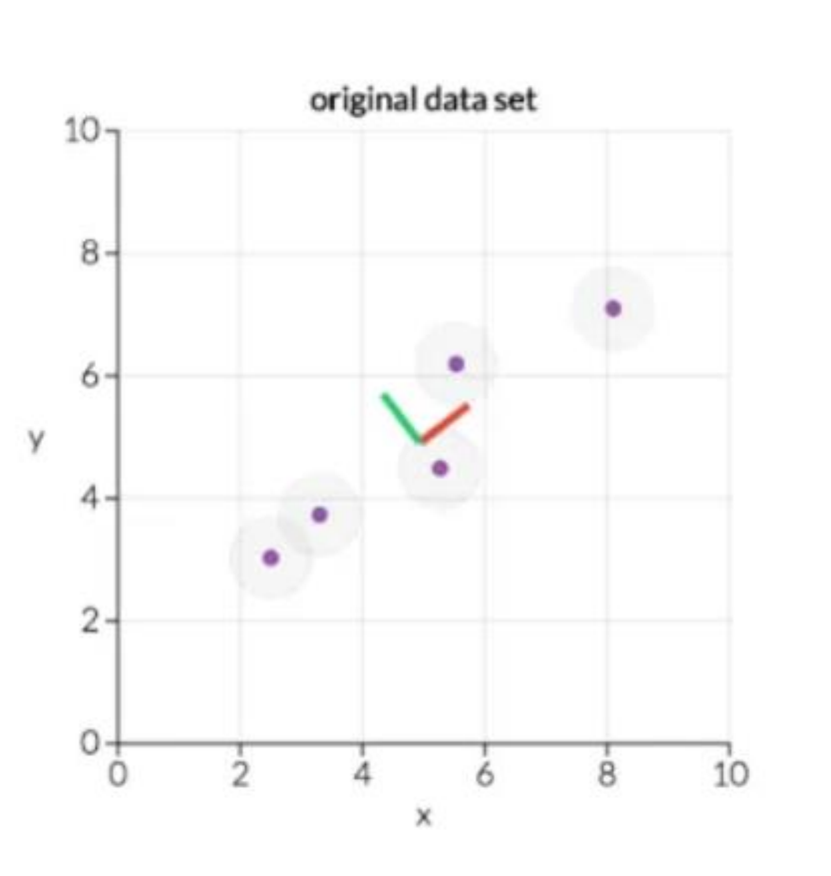
A continuación, descomponemos esta matriz en dos componentes separados: *dirección y magnitud*.

Así podremos comprender la dirección de nuestros datos y su magnitud

Análisis de Componentes Principales(PCA)

Se muestran las dos direcciones principales de estos datos: la dirección hacia atrás y la dirección verde. La dirección roja es la más importante

¿Por qué la dirección roja parece más importante que la verde?



Análisis de Componentes Principales(PCA)

¿Cómo sería la línea de mejor ajuste?

1. *Calcula la matriz de covarianza X de los puntos de datos*
2. *Calcule los vectores propios y los valores propios correspondientes*
3. *Ordenar los vectores propios según su valor dado en orden decreciente*
4. *Elegir los* k *primeros vectores propios y esas serán las nuevas* k *dimensiones*
5. *Transformar los puntos de datos originales* $n - dimensionales$ *en* $k\_dimensiones$

Análisis de Componentes Principales(PCA)

PCA es un método para encontrar la combinación lineal que tenga en cuenta la mayor variabilidad posible

$$PC = \alpha_1 \boxed{X_1} + \alpha_2 \boxed{X_2}$$

Por ejemplo, queremos combinar estas dos variables, PCA tratará de encontrar los valores óptimos para los pesos α_1 y α_2 de modo que las variables combinadas tengan la máxima varianza

Recordemos que los valore que usamos para los pesos son restringidos, así que la suma del cuadrado de los pesos es 1

$$\alpha_1{}^2 + \alpha_2{}^2 = 1$$

Análisis discriminante lineal(LDA)

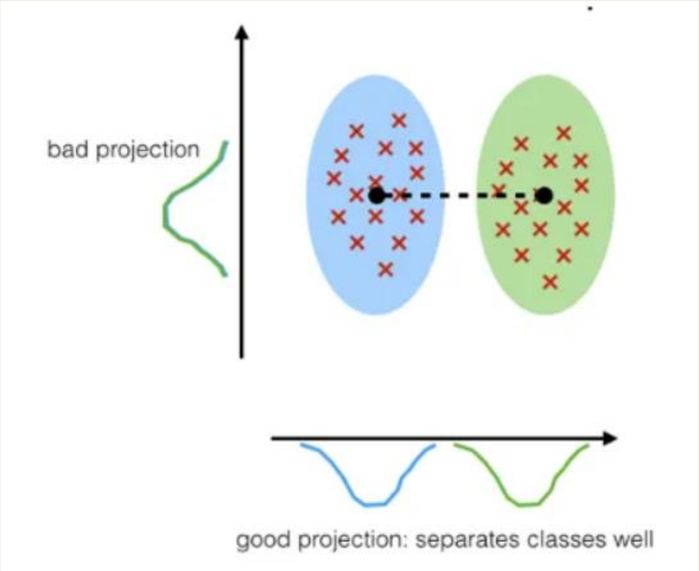
Lineal Discriminant Analysis (LDA)

- Es un tipo de combinación lineal, un proceso matemático que utiliza varios elementos de datos y les aplica una función para analizar por separado varias clases de objetos o elementos
- Siguiendo el discriminante lineal de Fisher, el análisis discriminante lineal puede ser útil en áreas como el reconocimiento de imágenes y el análisis predictivo en marketing
- Ayuda en los datos preventivos para más de dos clases, cuando la Regresión Logística no es suficiente

Análisis discriminante lineal(LDA)

Lineal Discriminant Analysis (LDA)

- El análisis discriminante lineal toma el valor medio de cada clase y considera variantes para hacer predicciones asumiendo una distribución gaussiana
- Maximización de los ejes componentes para la separación de clases.



Análisis discriminante lineal(LDA)

¿Cómo funciona el Análisis discriminante lineal?

1. *Calcular el vector medio* $d - dimensional$ *para las diferentes clases del conjunto de datos*
2. *Calcular la matriz de dispersión(entre clases y dentro de la matriz de dispersión de clase)*
3. *Ordenar el eigenvector por la disminución del eigenvalor y elegir el* k *eigenvector con el más grande eigenvalor a partir de una matriz* $d \times k$ *dimensional* w *(donde cada columna representa un eigenvector)*

Análisis discriminante lineal(LDA)

4. Se utiliza la matriz $d \times k$ eigenvector para transformar la muestra en el nuevo subespacio

Esto puede resumirse mediante la multiplicación de matrices

$$Y = X \times W$$

donde:

X: es una matriz de dimensión $n * d$ que representa las n muestras y se transforma $n * k$ muestras dimensionales en el nuevo subespacio

Análisis discriminante lineal(LDA)

Lineal Discriminant Analysis (LDA) es similar al método PCA, cuyo objetivo es maximizar la separación entre dos o más grupos

$$LDA = \alpha_1 \boxed{X_1} + \alpha_2 \boxed{X_2}$$

Por ejemplo, queremos combinar estas dos variables, LDA tratará de encontrar los valores óptimos para los pesos α_1 y α_2 de modo que las variables combinadas muestren la máxima separación entre el grupo

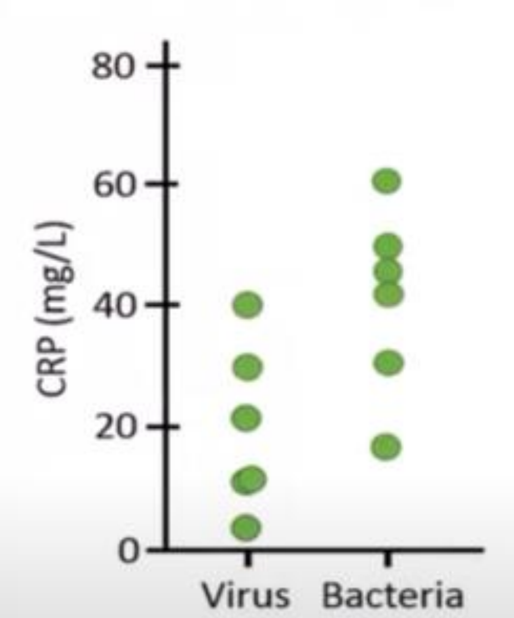
Separación

¿Esto que significa?

Infection	CRP (mg/L)	Temp (C)
Viral	40.0	36.0
Viral	11.1	37.2
Viral	30.0	36.5
Viral	21.4	39.4
Viral	10.7	39.6
Viral	3.4	40.7
Bacterial	42.0	37.6
Bacterial	31.1	42.2
Bacterial	50.0	38.5
Bacterial	60.4	39.4
Bacterial	45.7	38.6
Bacterial	17.3	42.7

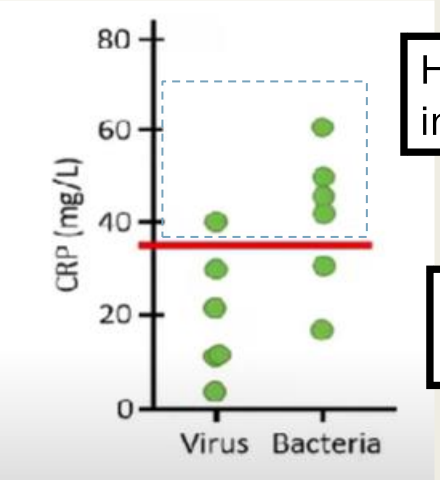
Concentración de la proteína c-reactive en la sangre

Si representamos gráficamente el CPR, podemos ver que las personas con infección virales suelen tener una menor concentración de CPR en comparación con las personas con infección bacterianas



Separación

Digamos que utilizaríamos un valor de corte de **35,** para determinar si alguien tiene una infección bacteriana o viral. Sin embargo, el problema de esta línea de corte es que no separa muy bien el grupo



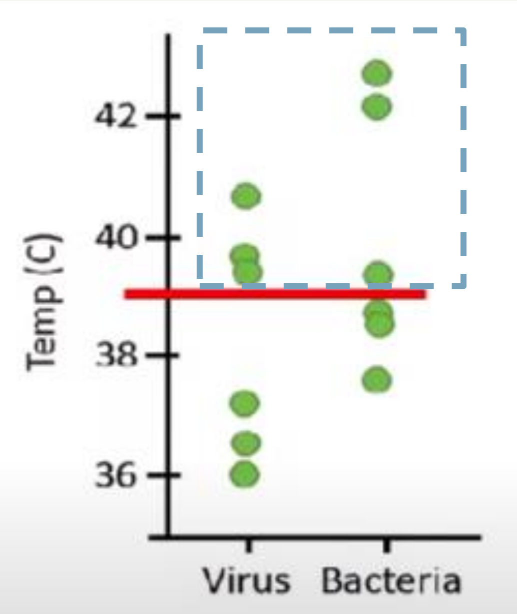
Hay 4 pacientes con infección bacteriana

Hay 2 pacientes con infección bacteriana

Usando esta línea o cualquier otra de manera horizontal será imposible separar estos dos grupos completamente

Separación

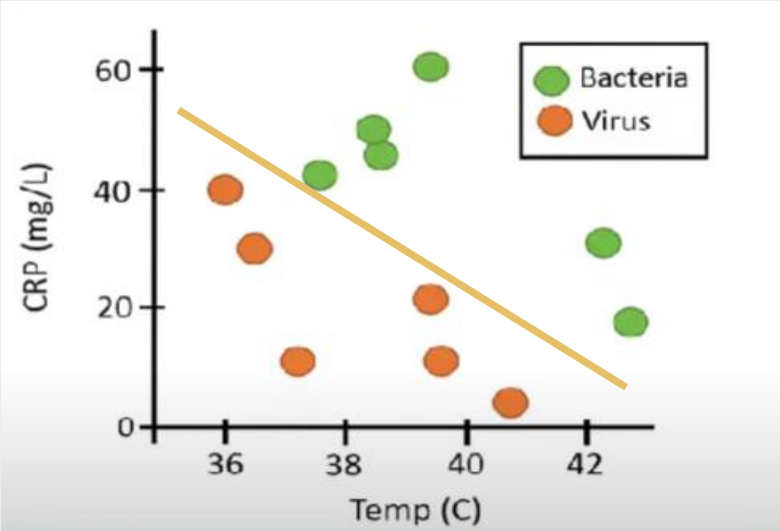
Si esta vez utilizamos la temperatura para graficar, otra vez tendremos el mismo problema



No hay línea que pueda separar en dos grupos a los pacientes

Separación

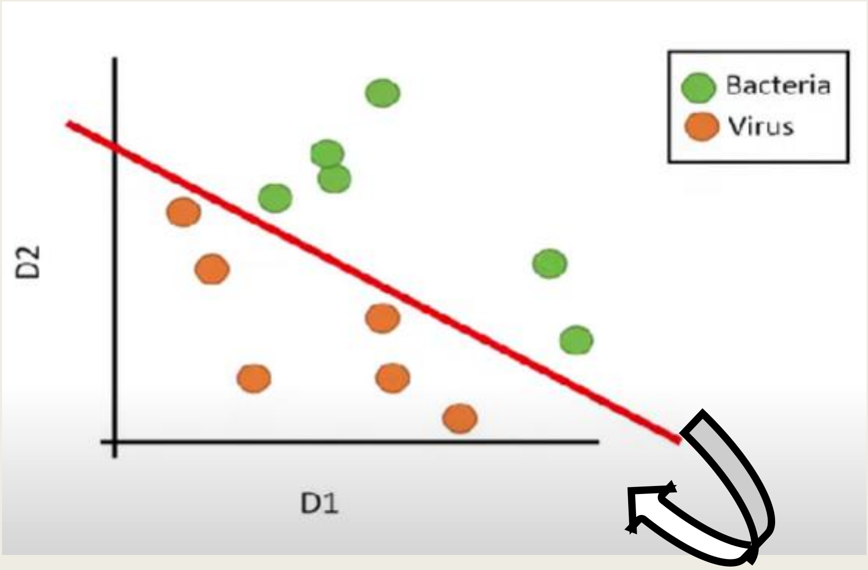
Si graficamos CPR y la temperatura y usamos diferentes colores para los pacientes:

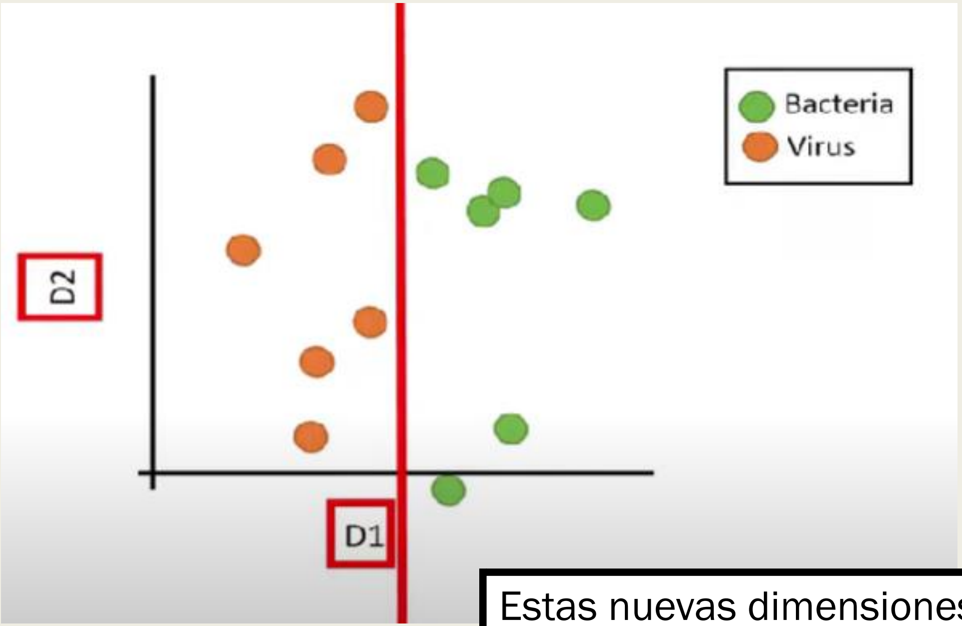


- Con esa línea podemos separar los grupos
- Al combinar ambas variables, podemos obtener mejores resultados que si solo usáramos una

Separación

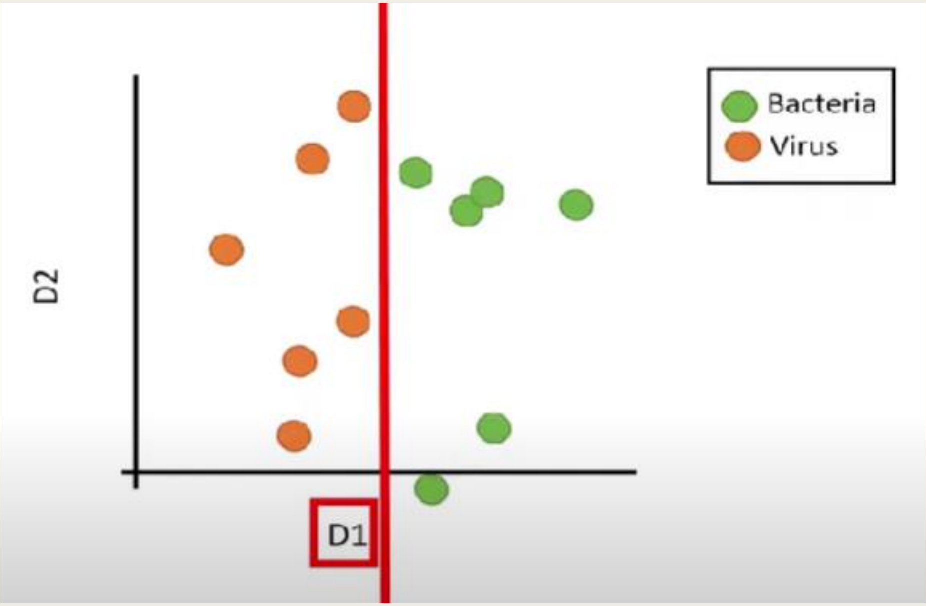
Justo PCA y LDA se pueden ver al girar los datos en dos nuevas dimensiones de la siguiente manera

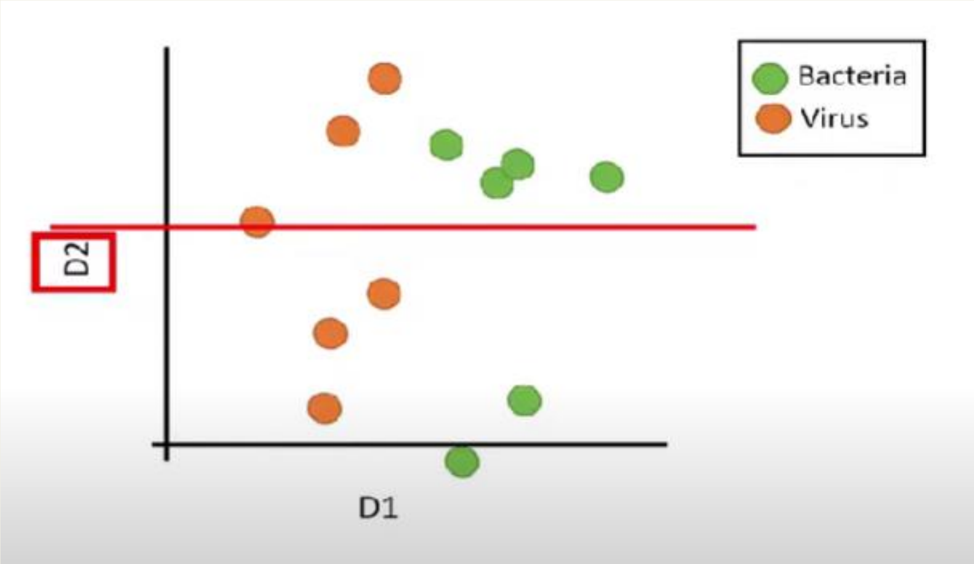




Estas nuevas dimensiones se llaman **función discriminante 1 y 2**

Separación

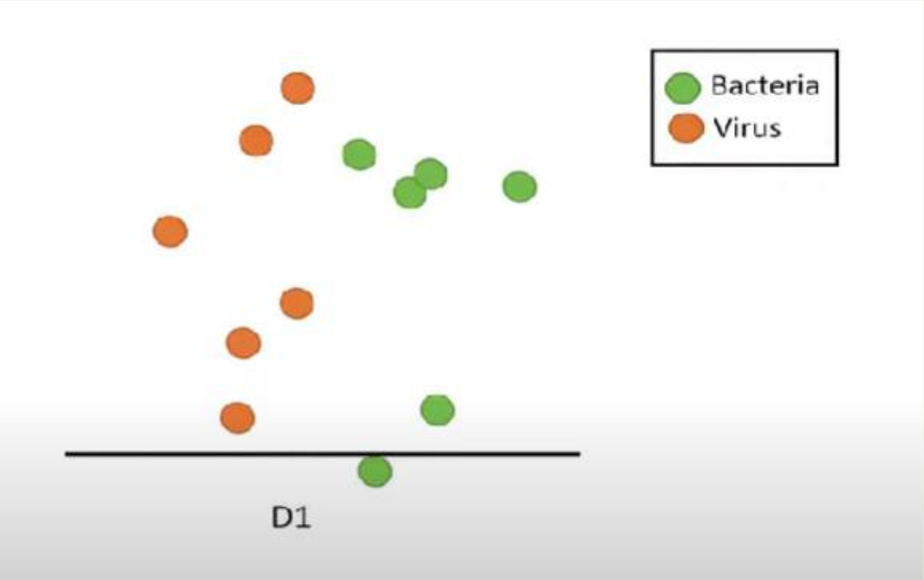


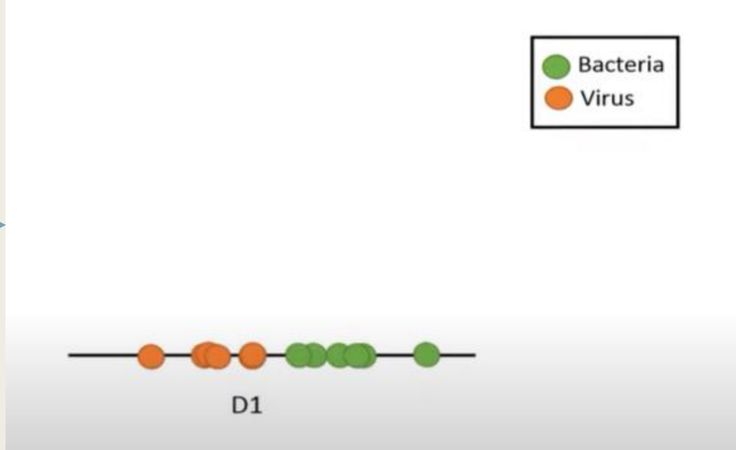


Podemos observar que la primera función discriminante (D1) separa perfectamente en dos grupos

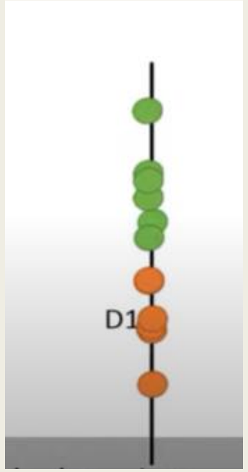
Pero, no podemos separar los grupos con una línea horizontal, esto significa que la segunda función discriminante(D2) no es un buen separador

Separación





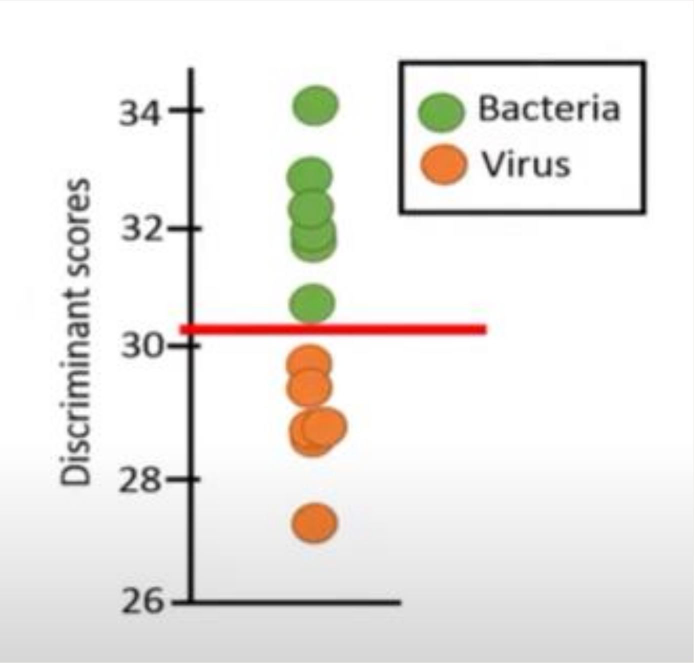
Podemos borrar D2 y colocar todos los datos en una sola línea



También podríamos colocarlo de manera horizontal

Análisis discriminante lineal(LDA)

Si usamos LDA, podemos graficar las llamadas puntuaciones discriminantes de la siguiente manera



- Podemos usar una línea horizontal que perfectamente nos separa los grupos
- Usando LDA, podemos combinar las variables CPR y temperatura para maximizar la separación entre los dos grupos

Análisis discriminante lineal(LDA)

Las puntuaciones discriminantes lineales que se ven en el gráfico podemos obtenerlas con la siguiente ecuación:

$$LDA = \alpha_1 X_1 + \alpha_2 X_2$$

Remplazando los valores

$$LDA = 0.11 \cdot CPR + 0.70 \cdot Temp$$

Estos valores nos dan la separación óptima entre los grupos

Usaremos estos valores para calcular la puntuación discriminante (discriminant scores)

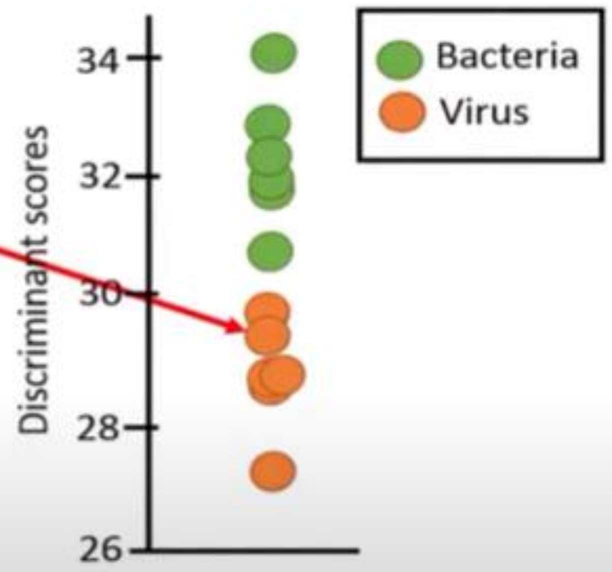
Análisis discriminante lineal(LDA)

$$\boxed{LDA = 0.11 \cdot CPR + 0.70 \cdot Temp}$$

$$LD = 0.11 \cdot 40.0 + 0.70 \cdot 36.0$$

Infection	CRP (mg/L)	Temp (C)	Scores
Viral	40.0	36.0	29.5
Viral	11.1	37.2	
Viral	30.0	36.5	
Viral	21.4	39.4	
Viral	10.7	39.6	
Viral	3.4	40.7	
Bacterial	42.0	37.6	
Bacterial	31.1	42.2	
Bacterial	50.0	38.5	
Bacterial	60.4	39.4	
Bacterial	45.7	38.6	
Bacterial	17.3	42.7	

representa

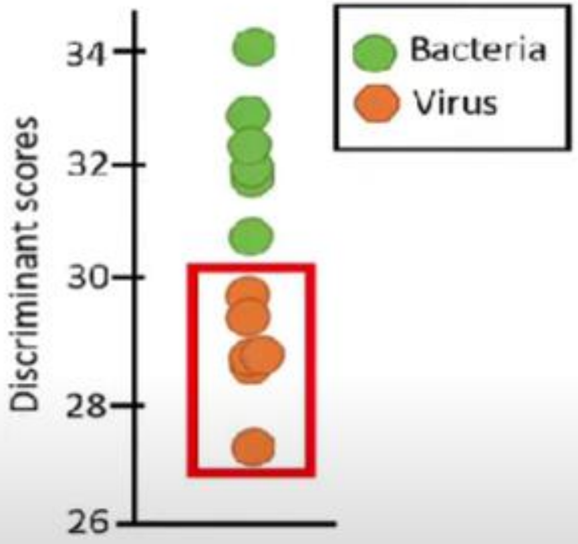


Análisis discriminante lineal(LDA)

Infection	CRP (mg/L)	Temp (C)	Scores
Viral	40.0	36.0	29.5
Viral	11.1	37.2	27.3
Viral	30.0	36.5	28.8
Viral	21.4	39.4	29.9
Viral	10.7	39.6	28.9
Viral	3.4	40.7	28.9
Bacterial	42.0	37.6	30.8
Bacterial	31.1	42.2	32.9
Bacterial	50.0	38.5	32.3
Bacterial	60.4	39.4	34.0
Bacterial	45.7	38.6	31.9
Bacterial	17.3	42.7	31.8

Menor valor

Mayor valor



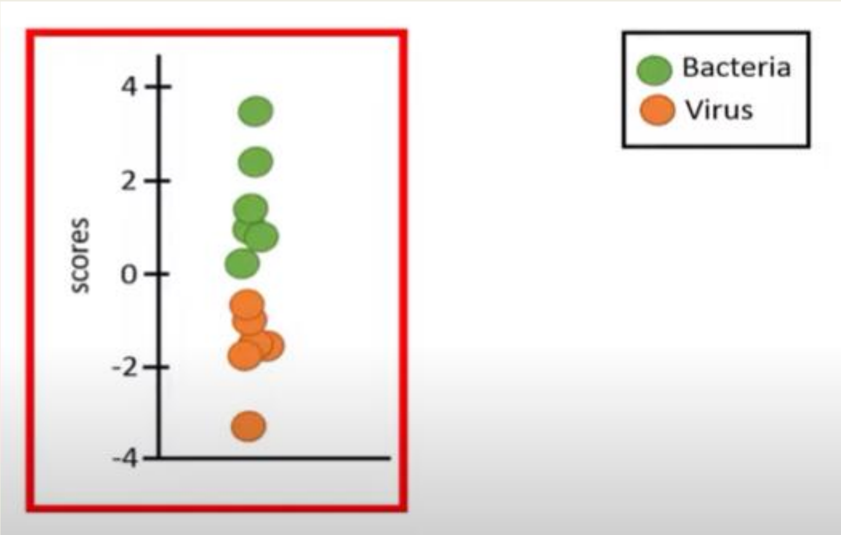
La mayoría de las herramientas informan de la puntuación discriminante centrada, debemos sustituir las medias correspondientes de los valores originales antes de calcular la puntuación

Análisis discriminante lineal(LDA)

Usando esta ecuación calculamos los centros

$$LD = 0.11 \cdot \left(\mathrm{CRP} - \overline{\mathrm{CRP}}\right) + 0.70 \cdot \left(\mathrm{Temp} - \overline{\mathrm{Temp}}\right)$$

Infection	CRP (mg/L)	Temp (C)	Scores	Cent. scores
Viral	40.0	36.0	29.5	-1.1
Viral	11.1	37.2	27.3	-3.3
Viral	30.0	36.5	28.8	-1.8
Viral	21.4	39.4	29.9	-0.7
Viral	10.7	39.6	28.9	-1.7
Viral	3.4	40.7	28.9	-1.7
Bacterial	42.0	37.6	30.8	0.2
Bacterial	31.1	42.2	32.9	2.3
Bacterial	50.0	38.5	32.3	1.7
Bacterial	60.4	39.4	34.0	3.5
Bacterial	45.7	38.6	31.9	1.3
Bacterial	17.3	42.7	31.8	1.2



Vemos que los scores centers están alrededor de 0, lo que significa que los scores son igual a 0

Discriminante lineal de Fisher

- Para empezar, consideremos el caso de un problema de clasificación de dos clases (K=2). Puntos azules y rojos en R^2

- En general, podemos tomar cualquier vector de entrada de $D-dimensiones$ y proyectarlo a $D' - dimensiones$

- Aquí, D representa las dimensiones de entrada originales, mientras que D' son las dimensiones del espacio proyectado. Consideraremos que $D' < D$

Discriminante lineal de Fisher

En el caso de la proyección a una dimensión (la recta numérica), es decir, $D' = 1$, podemos elegir un umbral t para separar las clases en el nuevo espacio. Dado un vector de entrada x:

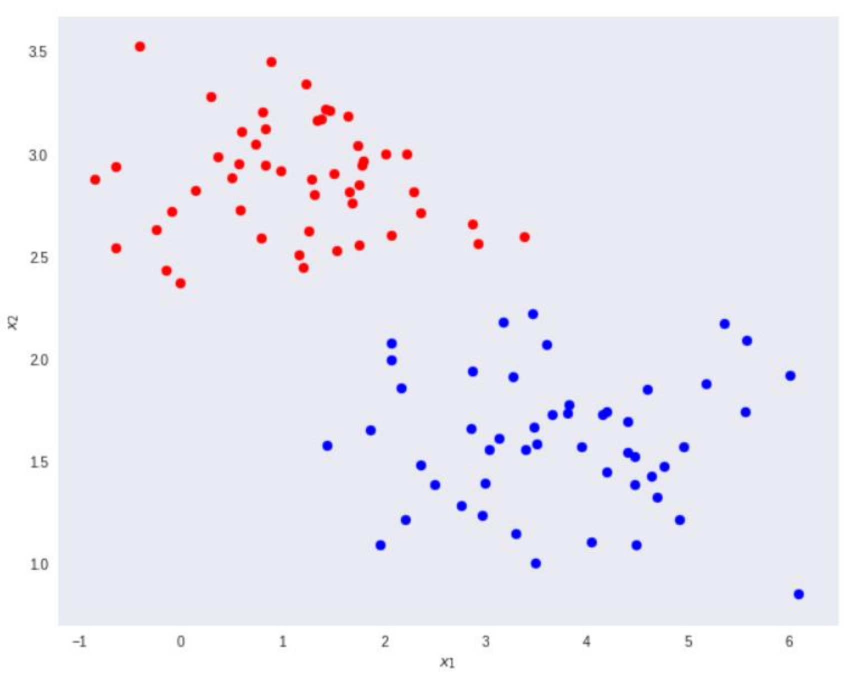
- *si el valor predicho $y >= t$ entonces, x pertenece a la clase C1 (clase 1) - donde $y = W^T x$*
- *en caso contrario, se clasifica como C2 (clase 2)*

Tomemos como ejemplo el siguiente dataset. Queremos reducir las dimensiones de los datos originales de $D = 2$ a $D' = 1$. En otras palabras, queremos una transformación T que mapee vectores en 2D a 1D –

$$T(v) = \mathbb{R}^2 \rightarrow \mathbb{R}^1$$

Discriminante lineal de Fisher

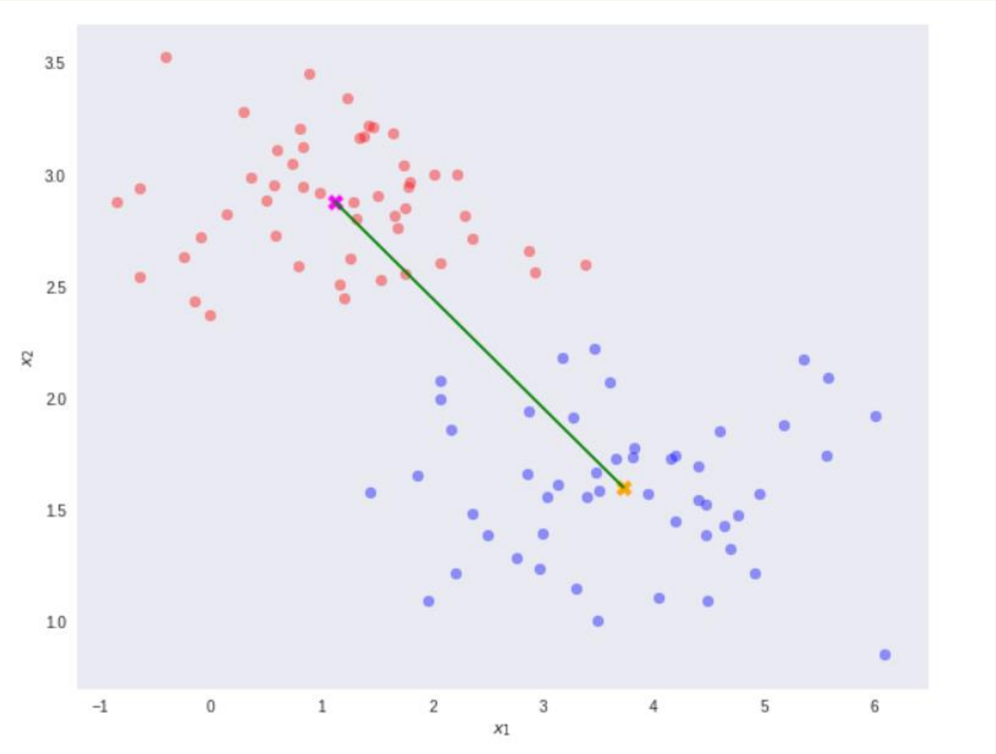
En primer lugar, vamos a calcular los vectores medios $\boldsymbol{m1}$ y $\boldsymbol{m2}$ para las dos clases



$$m_1 = \frac{1}{N_1} \sum_{n \in C1} x_n \qquad m_2 = \frac{1}{N_2} \sum_{n \in C2} x_n$$

- Nótese que $N1$ y $N2$ denotan el número de puntos en las clases $C1$ y $C2$ respectivamente
- Consideremos ahora el uso de las medias de clase como medida de separación
- En otras palabras, queremos proyectar los datos sobre el vector W que une las 2 medias de clase

Discriminante lineal de Fisher

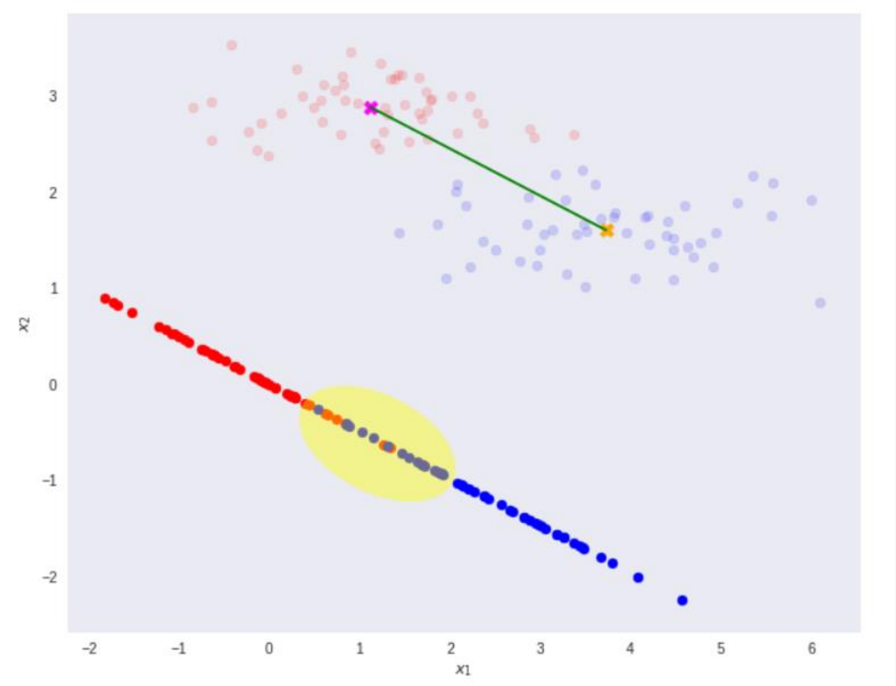


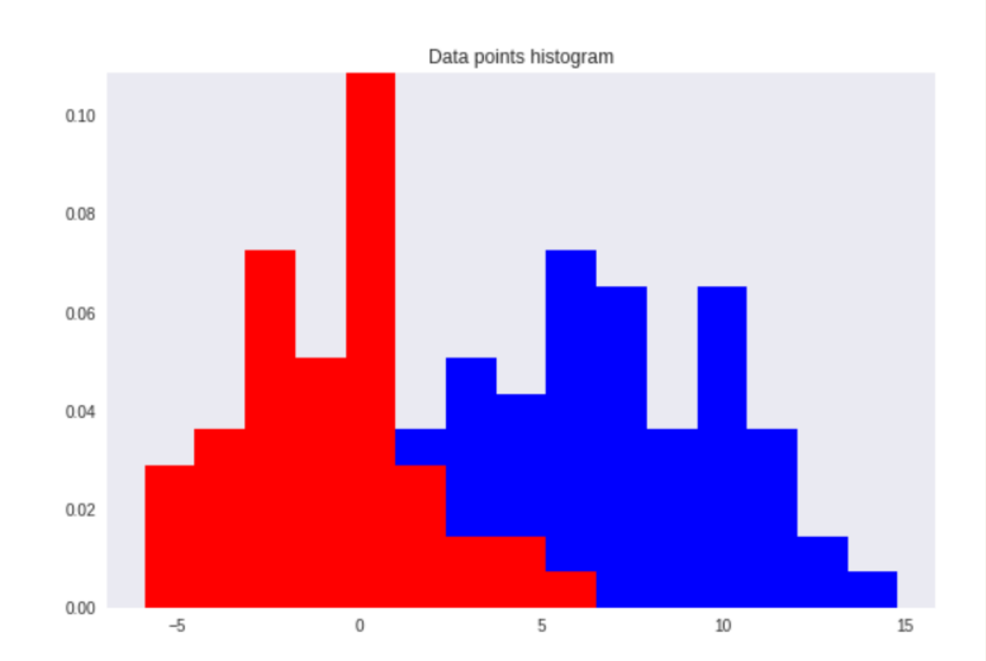
- Es importante señalar que cualquier tipo de proyección a una dimensión más pequeña puede implicar cierta pérdida de información

- En este caso, se observa que las dos clases son claramente separables (mediante una línea) en su espacio original

Discriminante lineal de Fisher

Sin embargo, tras la reproyección, los datos muestran una especie de solapamiento de clases, como muestran la elipse amarilla del gráfico y el histograma inferior





¿Qué es el solapamiento de clases? (overlapping clases)
Las muestras de datos aparecen como instancias válidas de más de una clase, lo que puede ser responsable de la presencia de ruido en los conjuntos de datos

Discriminante lineal de Fisher

Aquí es donde entra en juego el discriminante lineal de Fisher.

La idea propuesta por Fisher consiste en maximizar una función que proporcione una gran separación entre las medias de las clases proyectadas y, al mismo tiempo, una pequeña varianza dentro de cada clase, minimizando así el solapamiento de clases

Discriminante lineal de Fisher

En otras palabras, FLD selecciona una proyección que maximiza la separación de clases.

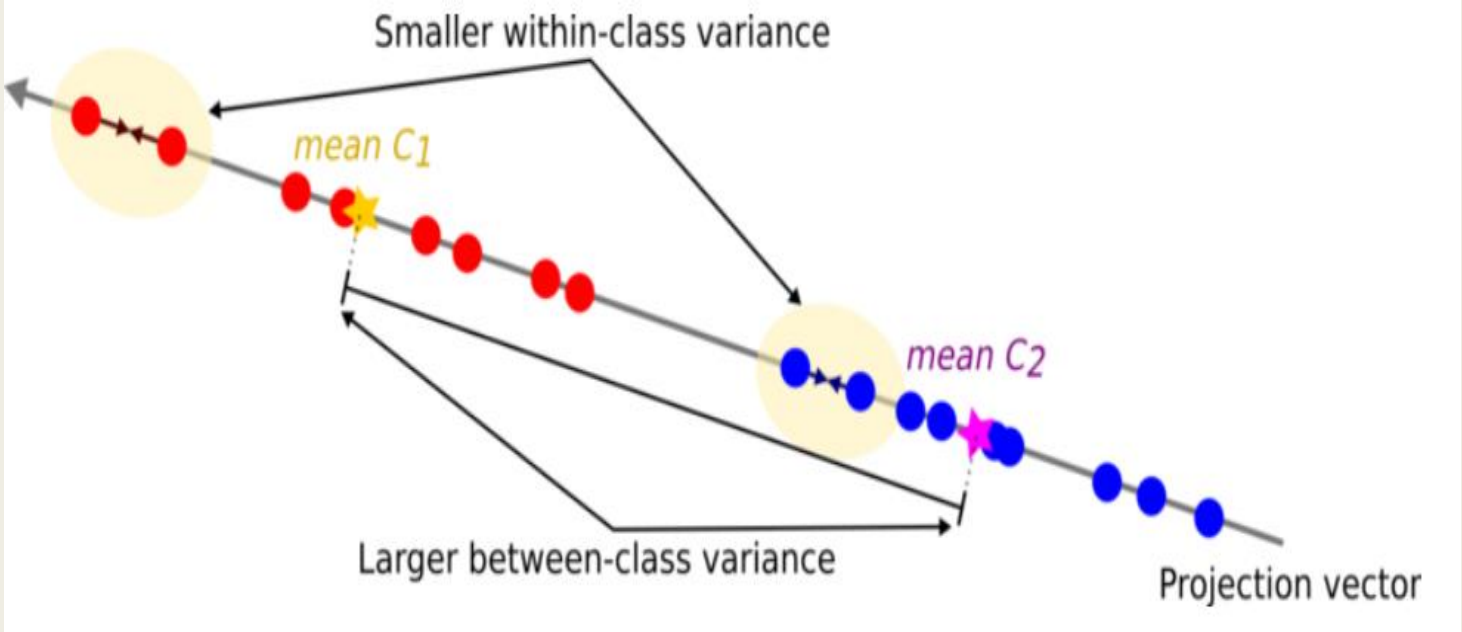
Para ello, maximiza la relación entre la varianza entre clases y la varianza dentro de clase.

En resumen, para proyectar los datos a una dimensión más pequeña y evitar el solapamiento de clases, FLD mantiene 2 propiedades.

- *Una gran varianza entre las clases del conjunto de datos*
- *Una varianza pequeña dentro de cada una de las clases del conjunto de datos*

Discriminante lineal de Fisher

Una pequeña varianza dentro de la clase tiene el efecto de mantener los puntos de datos proyectados más cerca unos de otros



Una gran varianza entre clases significa que las medias de las clases proyectadas deben estar lo más separadas posible

Discriminante lineal de Fisher

Para encontrar la proyección con las siguientes propiedades, FLD aprende un vector de pesos W con el siguiente criterio

$$J(\boldsymbol{W}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \qquad (1)$$

Between-class variance

Within-class variance

Si sustituimos los vectores medios $\boldsymbol{m1}$ y $\boldsymbol{m2}$, así como la varianza s, por las ecuaciones (1) y (2), llegamos a la ecuación (3)

Discriminante lineal de Fisher

$$s_k^2 = \sum_{n \in C_k} (y_n - m_k)^2 \qquad y_n = \boldsymbol{W}^T x_n \tag{2}$$

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T S_B \boldsymbol{W}}{\boldsymbol{W}^T S_W \boldsymbol{W}} \tag{3}$$

Si tomamos la derivada de (3) respecto a W (tras algunas simplificaciones) obtenemos la ecuación de aprendizaje para W (ecuación 4)

$$\boldsymbol{W} \propto S_W^{-1}(m_2 - m_1) \tag{4}$$

Per-class mean

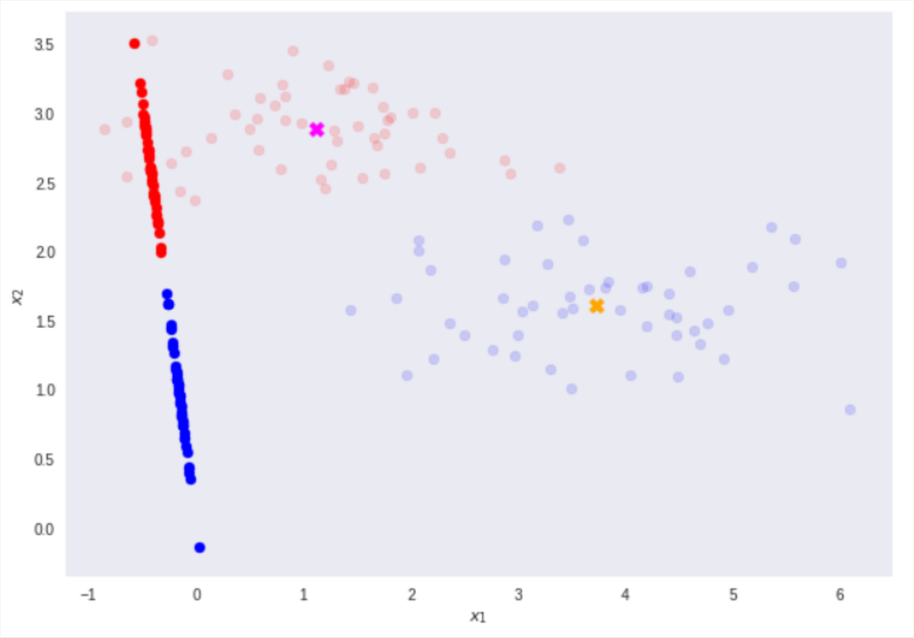
Within-class variance

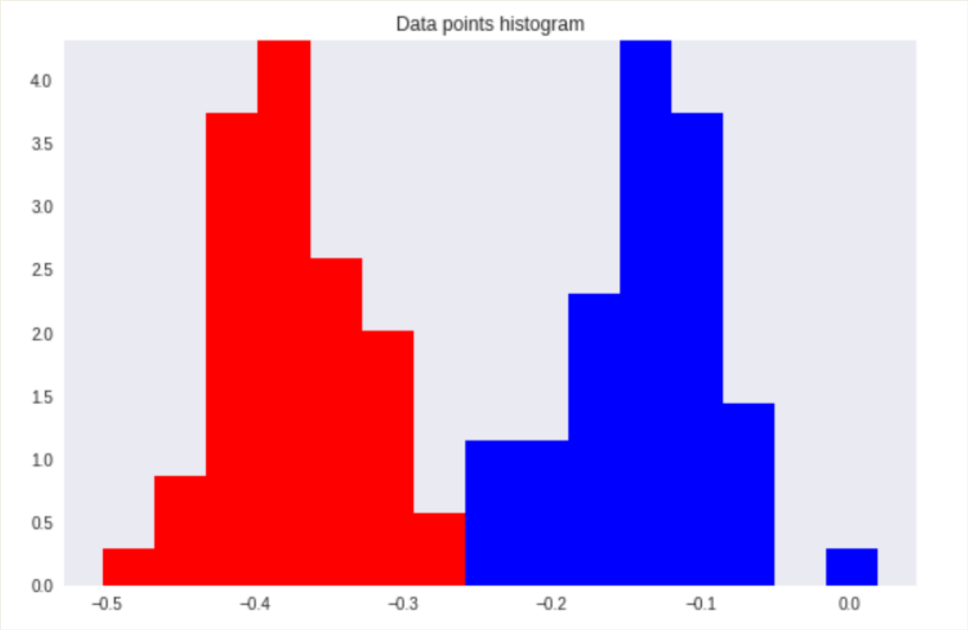
projection equation

Es decir, W (nuestra transformación deseada) es directamente proporcional a la inversa de la matriz de covarianza dentro de la clase multiplicada por la diferencia de las medias de las clases

Discriminante lineal de Fisher

El resultado permite una separación perfecta de las clases con un simple umbral





MATEMÁTICAS DE PCA

Análisis de Componentes Principales(PCA)

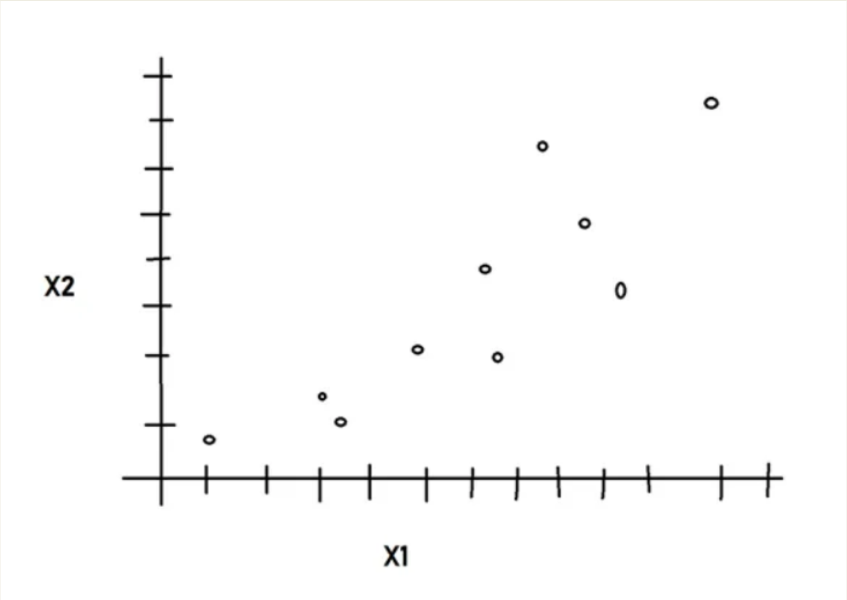
Ejemplo

DataSet:

X_1^*	X_2^*
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1.0	1.1
1.5	1.6
1.1	0.9

Tamaño $N = 10$
Variables $P = 2$

Construimos un diagrama de dispersión para ver cómo se distribuyen los datos



Análisis de Componentes Principales(PCA)

- **Su correlación:**

$$R^2 = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n(\sum y^2) - (\sum y)^2)}}$$

$$R = 0.926$$

Correlación positiva elevada
redundancia

- **Media de nuestras variables:**

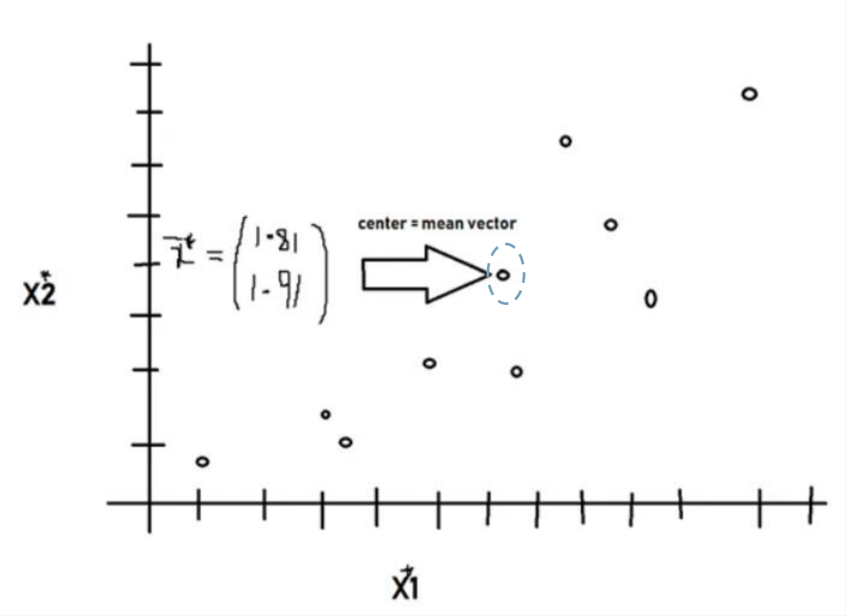
$$X_1^* = \frac{2.5+0.5+2.2.....+1.1}{10}$$

$$X_1^* = 1.81$$

$$X_2^* = \frac{2.4+0.7+2.9.....+0.9}{10}$$

$$X_2^* = 1.91$$

Entonces



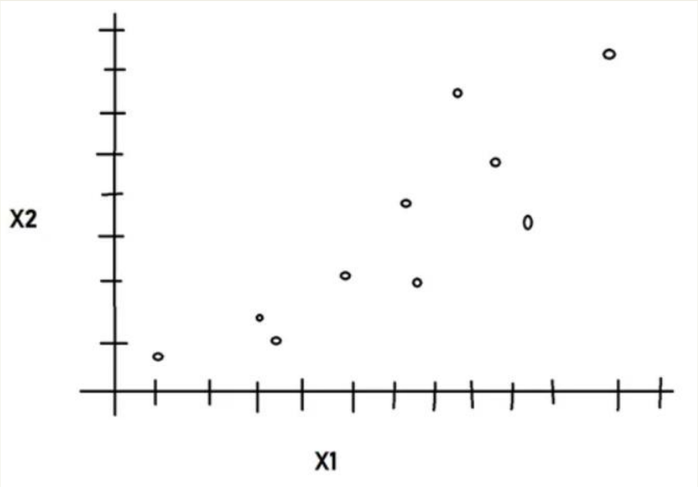
Análisis de Componentes Principales(PCA)

Paso 1.

- *Restar la media del componente de datos correspondiente para recentrar el conjunto de datos*
- *Reconstruir el diagrama de dispersión a la vista*
- *Escribir los datos "ajustados" como una matriz X.*

Nota: ese conjunto de datos "ajustado" tendrá media cero.

X_1^*	X_2^*	$X_1 = \overline{X_1^*} - X_1^*$	$X_2 = \overline{X_2^*} - X_2^*$
2.5	2.4	0.69	0.49
0.5	0.7	-1.31	-1.21
2.2	2.9	0.39	0.99
1.9	2.2	0.09	0.29
3.1	3.0	1.29	1.09
2.3	2.7	0.49	0.79
2.0	1.6	0.19	-0.31
1.0	1.1	-0.81	-0.81
1.5	1.6	-0.31	-0.31
1.1	0.9	-0.71	-1.01



Análisis de Componentes Principales(PCA)

Ahora escribimos los datos "ajustados" como una Matriz X.

$$X = \begin{pmatrix} 0.69 & 0.46 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix}_{10 \times 2}$$

Nota: el "conjunto de datos" ajustado tendrá media cero.

$$\overline{x} = \binom{0}{0}$$

Análisis de Componentes Principales(PCA)

Paso 2.

Calcular la matriz de varianza-covarianza de la muestra C.

$$C = \frac{1}{N-1}(X - 1\hat{X})(X - 1\hat{\bar{X}})$$

$$= \frac{1}{N-1}\hat{X}X$$

$$\hat{X} = transpose\ of\ X$$

$$C = \frac{1}{10-1}\begin{pmatrix} 0.69 & -1.31\ldots\ldots \\ 0.49 & -1.21\ldots\ldots \end{pmatrix} 2\ X\ 10\ \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21. \end{pmatrix} 10\ X\ 2$$

$$C = \frac{1}{10-1}\begin{pmatrix} 0.69 & -1.31\ldots\ldots \\ 0.49 & -1.21\ldots\ldots \end{pmatrix} 2\ X\ 10\ \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21. \end{pmatrix} 10\ X\ 2$$

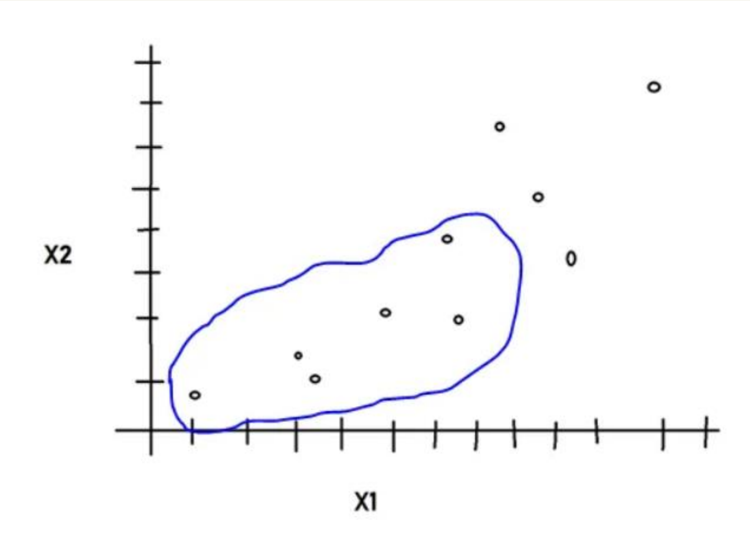
$$C = \frac{1}{9}\begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166. \end{pmatrix} 2\ X\ 2$$

Análisis de Componentes Principales(PCA)

Paso 3.

- *Calcular los eigenvalores lambda 1 y el eigenvalor*
- *De C : ordena los pares correspondientes de mayor a menor eigenvalor*

$$A = \begin{bmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{bmatrix}$$



Análisis de Componentes Principales(PCA)

$$\lambda I = \lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 0.6166 - \lambda & 0.6154 \\ 0.6154 & 0.7166 - \lambda \end{bmatrix}$$

$$= \det \begin{bmatrix} 0.6166 - \lambda & 0.6154 \\ 0.6154 & 0.7166 - \lambda \end{bmatrix}$$

$$=(0.61 - \lambda)(0.71 - \lambda) - (0.61)(0.6154)$$

$$= 0.441 - 0.661\,\lambda - 0.7166\,\lambda + \lambda^2 - 0.37$$

$$= \lambda^2 - 1.33\lambda + 0.0631$$

Aquí usamos la fórmula cuadrática para obtener los valore de $\lambda 1 \ y \ \lambda 2$:

$$\lambda 1 = 1.28 \ \text{and} \ \lambda 2 = 0.0492$$

These values are the eigenvalue

$$\text{Now} \begin{bmatrix} 0.6166 - \lambda & 0.6154 \\ 0.6154 & 0.7166 - \lambda \end{bmatrix}$$

$$\lambda 1 = 1.28$$

$$\text{So} \begin{bmatrix} 0.6166 - 1.28 & 0.6154 \\ 0.6154 & 0.7166 - 1.28 \end{bmatrix}$$

$$B = \begin{bmatrix} -0.6634 & 0.6154 \\ 0.6154 & -0.5634 \end{bmatrix}$$

$$\text{Solve } B\overline{X} = 0$$

$$\begin{bmatrix} -0.6634 & 0.6154 \\ 0.6154 & -0.5634 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Análisis de Componentes Principales(PCA)

Después de resolver la matriz, obtenemos el valor del eigenvector

x1= 0.678 and x2=0.735

Eigen Vector $\begin{bmatrix} 0.678 \\ 0.735 \end{bmatrix}$

Similarly for $\lambda 2 = 0.0492$

So Eigen Vector $\begin{bmatrix} 0.735 \\ -0.678 \end{bmatrix}$

Por lo que tenemos los valores de ambos: eigenvalores y eigenvectores

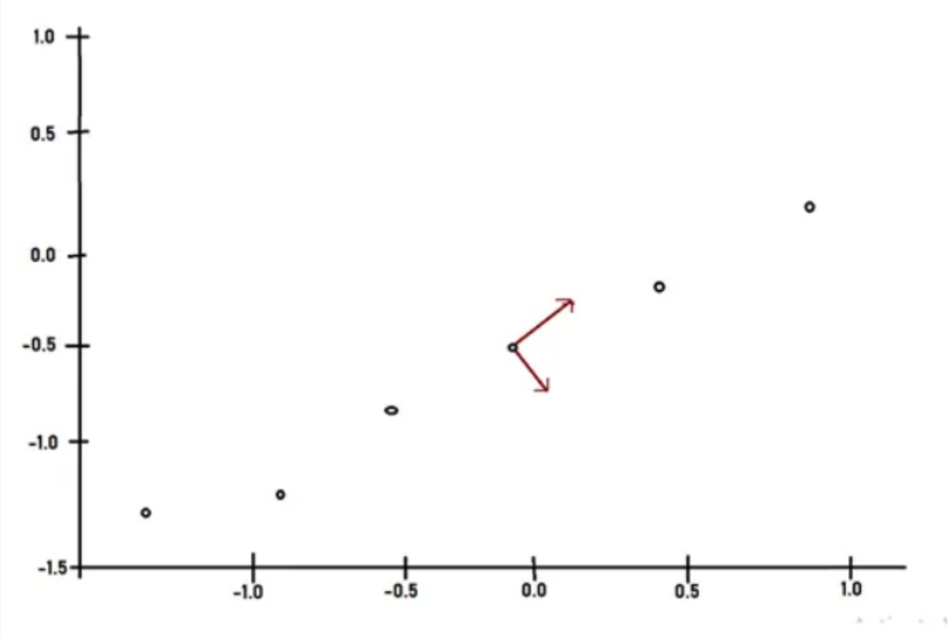
Variable	Eigen Vector 1	Eigen Vector 2
X1	0.678	0.735
X2	0,735	-0.678
Eigen Value	$\lambda 1 = 1.28$	$\lambda 1 = 0.49$
% total Variance	$\frac{1.28}{1.33} = 96\%$ PC 1	$\frac{0.0490}{1.33} = 3.7\%$ PC 2

Análisis de Componentes Principales(PCA)

El eigenvector:

$$\lambda 1 = \begin{pmatrix} 0.678 \\ 0.735 \end{pmatrix}$$

$$\lambda 2 = \begin{pmatrix} 0.735 \\ -0.678 \end{pmatrix}$$



En Eigen Vector1 se mueve a la derecha y 0.735 direcciones hacia arriba
En Eigen Vector2 se mueve a la derecha y -0.678 direcciones hacia arriba

Análisis de Componentes Principales(PCA)

Se puede demostrar:

$$Varianza\ total\ de\ la\ muestra = Suma\ del\ eigen\, Valor$$
$$= 1.28 + 0.0429 = 1.33$$

Con este proceso, podremos obtener líneas exactas que caractericen los datos. El primer eigenvector pasará por el medio de los puntos de datos como si fuera la línea de mejor ajuste.

El segundo eigenvector nos dará el otro patrón menos importante de los datos. Es decir, todos los puntos de datos siguen la línea principal. ¿Pero están desviados de la línea principal por alguna cantidad?

Análisis de Componentes Principales(PCA)

Paso 4.

- *Elijen los componentes y se forma la matriz de eigenvector V. Al ordenar los eigenvectores acorde con el eigenvalor, se obtienen los componentes por orden de significancia*
- *Por lo tanto, el eigenvector más alto es el componente principal. Los componentes menos significativos pueden ignorarse para reducir las dimensiones del conjunto de datos*

Análisis de Componentes Principales(PCA)

Seleccionamos ambos componentes:

$$V = \begin{pmatrix} 0.678 & 0.678 \\ 0.735 & 0.735 \end{pmatrix}$$

Descartamos el componente menos significativo (tomamos PC1 que tiene un 96% de captura de información)

$$V = \begin{pmatrix} 0.678 \\ 0.735 \end{pmatrix}$$

Análisis de Componentes Principales(PCA)

Paso 5.

- *Dividimos el nuevo dataset tomando* $Y = XV$

Hemos transformado nuestros datos. De modo que se expresan en términos del patrón entre ellos, donde los patrones son las líneas que describen más estrechamente la relación entre los datos.

$$Y = \begin{array}{cc} \left[\begin{array}{cc} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ \vdots & \vdots \\ 1.1 & 0.9 \end{array}\right] & \left[\begin{array}{cc} 0.678 & 0.735 \\ 0.735 & -0.678 \end{array}\right] \\ \boxed{X_1^*} \quad \boxed{X_2^*} & \boxed{V} \end{array}$$

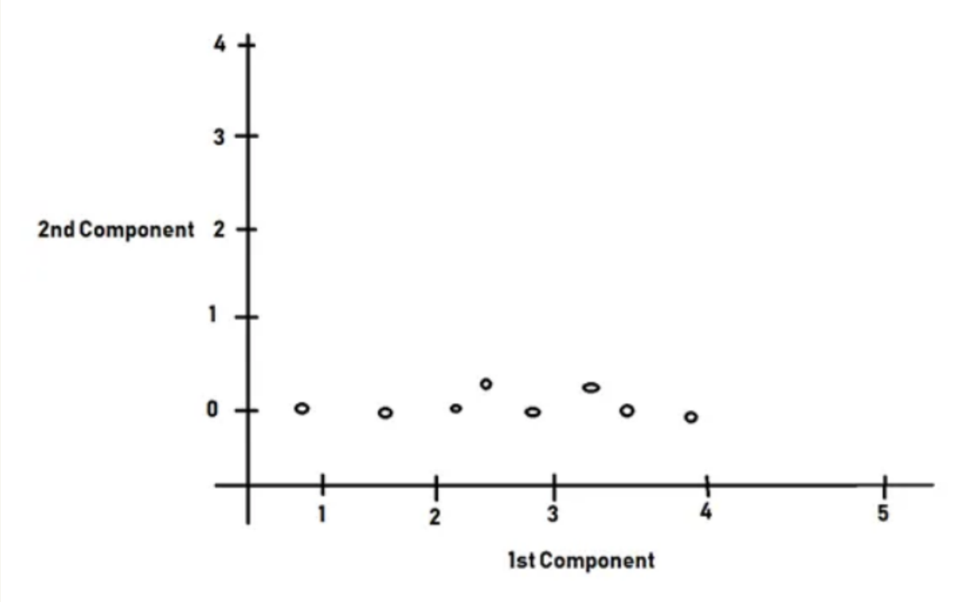
$$Y = \left[\begin{array}{cc} 3.459 & 0.211 \\ 0.854 & -0.107 \\ \vdots & \vdots \\ 1.407 & 0.199 \end{array}\right]$$

Análisis de Componentes Principales(PCA)

$$Y1 = 0.678\, x_1^* + 0.735\, x_2^*$$

$$Y1 = 0.735\, x_1^* - 0.678\, x_2^*$$

Score plot $x_1^* \ldots\ldots x_2^*$



Ahora, descartamos de nuevo el componente significativo

$$V = \begin{pmatrix} 0.738 \\ -0.678 \end{pmatrix}$$

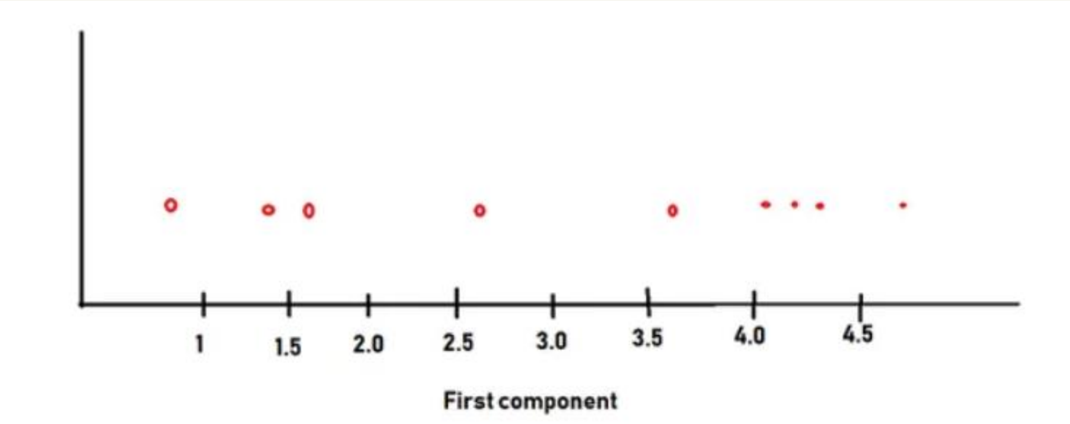
$$y = XV$$

Análisis de Componentes Principales(PCA)

$$Y = \begin{bmatrix} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1.1 & 0.9 \end{bmatrix} \begin{bmatrix} 0.678 \\ \\ 0.735 \end{bmatrix}$$

$\boxed{X_1^*}$ $\boxed{X_2^*}$ $\boxed{V}$

$$Y = \begin{bmatrix} 3.459 \\ 0.854 \\ \cdot \\ \cdot \\ \cdot \\ 1.407 \end{bmatrix}$$



En este caso, el PCA reduce una dimensión

$$\boxed{Y1 = 0.678\, x_1^* + 0.735\, x_2^*}$$

MATEMÁTICAS DE LDA

Análisis discriminante lineal(LDA)

Ejemplo

Dataset:

$$W1 = \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3\\5\end{pmatrix}, \begin{pmatrix}4\\3\end{pmatrix}, \begin{pmatrix}5\\6\end{pmatrix}, \begin{pmatrix}7\\5\end{pmatrix}$$

$$W1 = \begin{pmatrix}6\\2\end{pmatrix}, \begin{pmatrix}9\\4\end{pmatrix}, \begin{pmatrix}10\\1\end{pmatrix}, \begin{pmatrix}12\\3\end{pmatrix}, \begin{pmatrix}13\\6\end{pmatrix}$$

Aquí $W1$ y $W2$ son dos clases diferentes $W1$ pertenece a la clase 1 y $W2$ pertenece a la clase 2

Para la clase 1:

W1				
1	3	4	5	7
1	5	3	6	5

$$\mu1 = \frac{1+3+4+5+7+6+9+10+12+13}{10}$$

$$\mu1 = 3.5$$

Para la clase 2:

W2				
6	9	10	12	13
2	4	1	3	6

$$\mu2 = \frac{1+5+3+6+5+2+4+1+3+6}{10}$$

$$\mu2 = 3.7$$

Análisis discriminante lineal(LDA)

x1	x2	x1- μ1	x2- μ2
1	2	-2.5	-0.3
3	5	−1.7	1.3
4	3	-0.5	-0.7
5	6	1.5	2.3
7	5	3.5	1.3
6	2	2.5	-1.7
9	4	5.5	0.3
10	1	6.5	-2.7
12	3	8.5	-0.7
13	6	9.5	2.3

Ahora

$$\sum = \mathrm{E}\,\{(x - \mu)(\{(x - \mu)^t\}\}$$

Se aplica a la tabla anterior y obtenemos

$$\text{For Row 1} > \begin{pmatrix} -2.5 \\ -0.3 \end{pmatrix}(-2.5 \;\; -0.3) = \begin{pmatrix} 0.625 & 0.75 \\ 0.75 & 0.09 \end{pmatrix}$$

Esta es la Métrica 1

Análisis discriminante lineal(LDA)

Del mismo modo, el procedimiento para otras filas

$$M2 = \begin{pmatrix} 0.25 & -0.65 \\ -0.65 & 1.69 \end{pmatrix}$$

$$M3 = \begin{pmatrix} 0.25 & -0.35 \\ -0.35 & 0.49 \end{pmatrix}$$

$$M4 = \begin{pmatrix} 2.25 & 3.45 \\ 3.45 & 5.29 \end{pmatrix}$$

$$M5 = \begin{pmatrix} 12.25 & 4.55 \\ 4.55 & 1.69 \end{pmatrix}$$

$$M6 = \begin{pmatrix} 6.5 & -4.25 \\ -4.25 & 2.89 \end{pmatrix}$$

$$M7 = \begin{pmatrix} 30.25 & 1.65 \\ 1.65 & 0.09 \end{pmatrix}$$

$$M8 = \begin{pmatrix} 42.25 & -17.55 \\ -17.55 & 7.29 \end{pmatrix}$$

$$M9 = \begin{pmatrix} 72.25 & -5.95 \\ -5.95 & 0.49 \end{pmatrix}$$

$$M10 = \begin{pmatrix} 90.25 & 21.85 \\ 21.85 & 5.29 \end{pmatrix}$$

Ahora

$$\Sigma = \mathrm{E}\,\{(x - \mu)(\{(x - \mu)^t\}\}$$

$$= \frac{1}{10} \sum_{i=1}^{10} (x - \mu)(x - \mu)^t$$

$$= \frac{1}{10} \sum_{i=1}^{10} Mi$$

$$= \begin{pmatrix} 26.275 & 0.35 \\ 0.35 & 2.53 \end{pmatrix}$$

esta es la matriz de covarianza

Análisis discriminante lineal(LDA)

Encontramos el eigenvalor y la eigen matriz

$$\begin{bmatrix} 26.275-\lambda & 0.35 \\ 0.35 & 2.53-\lambda \end{bmatrix} = 0$$

$$(26.275-\lambda)(2.53-\lambda) - (0.35)^2 = 0$$

$$\lambda^2 - 28.805\,\lambda + 66.3532 = 0$$

Aplicamos la fórmula cuadrática y obtenemos los 2 valores lambda, estos valores son los eigenvalores

$\lambda 1 = 26.28 \quad , \lambda 2 = 2.525$

$$AX = \lambda X$$

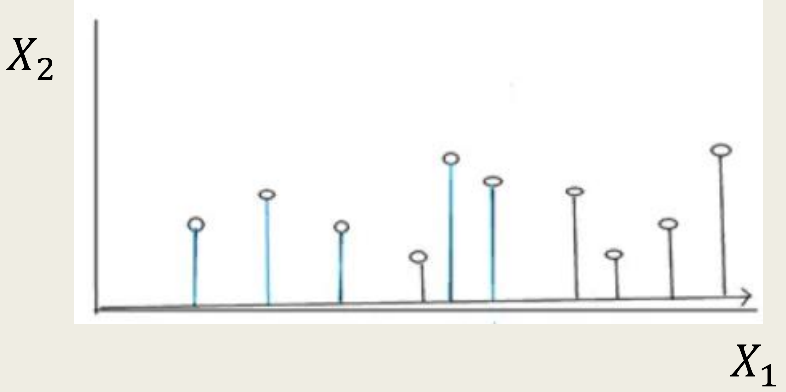
$$\begin{pmatrix} 26.275 & 0.35 \\ 0.35 & 2.53 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \end{pmatrix} = 26.28 \begin{pmatrix} x1 \\ x2 \end{pmatrix}$$

$26.275\,x1 + 0.35x2 = 26.028\,x1$

$0.35\,x1 + 2.53\,x2 = 26.28\,x2$

$X1 = 70,\ X2 = 1$

$$\begin{pmatrix} 70 \\ 1 \end{pmatrix}$$



$Sw = S1 + S2$

$W = s^{-1}w\,(\mu 1 - \mu 2)$

Análisis discriminante lineal(LDA)

W1	1	3	4	5	7	**μ1 = 4**
	2	5	3	6	5	**= 4.2**
W2	6	9	10	12	13	**μ2 = 10**
	2	4	1	3	6	**= 3.5**

$$S1 = \sum (xi - \mu 1)(xi \ - \mu 1)^t$$

$$S2 = \sum (xi - \mu 2)(xi \ - \mu 2)^t$$

Ahora para $w1$

X - μ1	-3	-1	0	1	3
	-2.2	0.8	-1.2	1.8	0.8

Análisis discriminante lineal(LDA)

Ahora para $w2$

X – μ2	-4	-1	0	2	3
	-1.2	0.8	-2.2	-0.2	2.8

Now $(x-\mu1))\ (x-\mu1)^t$ From w1

$$\begin{pmatrix}-3\\-2.2\end{pmatrix}(-3\quad -2.2)=\begin{pmatrix}9 & 6.6\\6.6 & 4.84\end{pmatrix}$$

$$\begin{pmatrix}-1\\0.8\end{pmatrix}(-1\quad -0.8)=\begin{pmatrix}1 & -0.8\\-0.8 & 0.64\end{pmatrix}$$

$$\begin{pmatrix}0\\1.2\end{pmatrix}(0\quad 1.2)=\begin{pmatrix}0 & 0\\0 & 1.44\end{pmatrix}$$

$$\begin{pmatrix}1\\1.8\end{pmatrix}(1\quad 1.8)=\begin{pmatrix}1 & 1.8\\1.8 & 3.24\end{pmatrix}$$

$$\begin{pmatrix}3\\0.8\end{pmatrix}(3\quad 0.8)=\begin{pmatrix}9 & 2.4\\2.4 & 0.64\end{pmatrix}$$

Sum of 5 matrices (w1)

$$S1=\begin{pmatrix}9 & 6.6\\6.6 & 4.84\end{pmatrix}+\begin{pmatrix}1 & -0.8\\-0.8 & 0.64\end{pmatrix}+\begin{pmatrix}0 & 0\\0 & 1.44\end{pmatrix}+\begin{pmatrix}1 & 1.8\\1.8 & 3.24\end{pmatrix}+\begin{pmatrix}9 & 2.4\\2.4 & 0.64\end{pmatrix}$$

$$S1=\begin{pmatrix}20 & 10\\10 & 10.8\end{pmatrix}$$

Now $(x-\mu1))\ (x-\mu1)^t$ From w2

$$\begin{pmatrix}-4\\-1.2\end{pmatrix}(-4\quad -1.2)=\begin{pmatrix}16 & 4.8\\4.8 & 1.44\end{pmatrix}$$

$$\begin{pmatrix}-1\\0.8\end{pmatrix}(-1\quad -0.8)=\begin{pmatrix}1 & -0.8\\-0.8 & 0.64\end{pmatrix}$$

$$\begin{pmatrix}0\\-2.2\end{pmatrix}(0\quad -2.2)=\begin{pmatrix}0 & 0\\0 & 4.84\end{pmatrix}$$

$$\begin{pmatrix}2\\-0.2\end{pmatrix}(2\quad -0.2)=\begin{pmatrix}4 & -0.4\\-0.4 & 0.04\end{pmatrix}$$

$$\begin{pmatrix}3\\2.8\end{pmatrix}(3\quad 2.8)=\begin{pmatrix}9 & 8.4\\8.4 & 7.84\end{pmatrix}$$

Sum of 5 matrices (w2)

$$S2=\begin{pmatrix}16 & 4.8\\4.8 & 1.44\end{pmatrix}+\begin{pmatrix}1 & -0.8\\-0.8 & 0.64\end{pmatrix}+\begin{pmatrix}0 & 0\\0 & 4.84\end{pmatrix}+\begin{pmatrix}4 & -0.4\\-0.4 & 0.04\end{pmatrix}+\begin{pmatrix}9 & 8.4\\8.4 & 7.84\end{pmatrix}$$

$$S2=\begin{pmatrix}30 & 12\\12 & 14.8\end{pmatrix}$$

Análisis discriminante lineal(LDA)

Después de encontrar el S1 y S2 podemos Sw

$S_w = S1 + S2$

$S_w = \begin{pmatrix} 20 & 10 \\ 10 & 10.8 \end{pmatrix} + \begin{pmatrix} 30 & 12 \\ 12 & 14.8 \end{pmatrix}$

Now find S_w^{-1} =?

$S_w^{-1} = \frac{1}{det\begin{pmatrix} 50 & 22 \\ 22 & 25.6 \end{pmatrix}} \begin{pmatrix} 25.6 & 22 \\ -22 & 50 \end{pmatrix}$

$S_w^{-1} = \begin{pmatrix} 0.032 & -0.03 \\ -0.03 & 0.06 \end{pmatrix}$

Now $\vec{e} = S_w^{-1}\ (\mu 1 - \mu 2)$

$\vec{e} = \begin{pmatrix} 0.032 & -0.03 \\ -0.03 & 0.06 \end{pmatrix} \left(\begin{pmatrix} 4 \\ 4.2 \end{pmatrix} - \begin{pmatrix} 10 \\ 3.2 \end{pmatrix} \right)$

$\vec{e} = \begin{pmatrix} -0.222 \\ 0.24 \end{pmatrix}$