

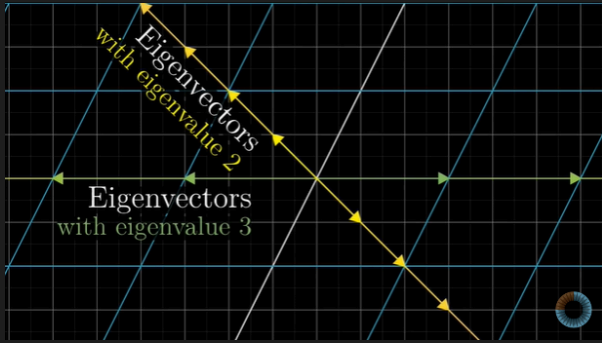
ÁLGEBRA

Autovectores

- son los vectores que no ven afectada su dirección cuando se aplica una transformación al espacio vectorial.
- Solo se ven escalados por un factor
- en un espacio 3D, si encuentro un autovector en realidad estoy encontrando el eje de rotación de la transformación

Autovalores

- el factor por el cual son escalados los autovectores



Hay que encontrar los valores para los cuales esta expresión es cierta:

$$\underbrace{A\vec{v}}_{\text{mult. entre matriz y vector}} = \underbrace{\lambda\vec{v}}_{\text{mult. entre escalar y vector}}$$

$$(A - \lambda I)\vec{v} = \vec{0}$$

we want a nonzero solution for \vec{v}

La única forma para que esto sea cierto, es que la transformación asociada a esa matriz $(A - \lambda I)$ aplaste el espacio a una dimensión menor. Ese aplastamiento corresponde a $\det(A - \lambda I) = 0$

"Una rotación en el plano no tiene autovectores"

$$\det\left(\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}\right) = \lambda^2 + 1 = 0 \Rightarrow \lambda_{1,2} = \pm i$$

El hecho de que no hayan soluciones reales, indica que no hay autovectores

Siempre que una matriz está llena de ceros, excepto en su diagonal principal, la llamamos "matriz diagonal"

- todas las columnas de la matriz son autovectores
- todos los elementos de la diagonal son autovalores

$\Lambda^{-1}A\Lambda$ es una matriz diagonal

PROBABILIDAD

Variable aleatoria

- formas de mapear los resultados de procesos aleatorios a números

Random process

Outcomes \rightarrow numbers (estamos cuantificando las salidas)

Coin flip: $X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$

Y = suma de las caras hacia arriba después de tirar 7 dados

Why are we doing this? Why is this useful?

- as soon as you start quantify outcomes, you can start to do a little bit more math on the outcomes, and you can start to use a little bit more mathematical notation on the outcome.

Las variables aleatorias pueden ser

- **Discretas**
 - valores separados
 - coin flip X
 - Y = year that a random student was born
 - Z = # de hormigas que van a nacer mañana en el universo
 - X = winning time for the men's 100 m dash 2016 Olympics, rounded to the nearest hundredth
- **Continuas**
 - pueden obtener cualquier valor en un rango. El rango incluso puede ser infinito
 - Y = exact mass of a random animal selected at the New Orleans zoo
 - X = exact winning time for the men's 100 m dash 2016 Olympics

INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO NO SUPERVISADO

- Una IA aprenda algo en base a datos.
- Clustering
 - ppal técnica del aprendizaje automático no supervisado

CRONOGRAMA

Clase 1 - 02/05/2023. Teoría y práctica de clustering

Clase 2 - 05/05/2023. Teoría de reducción de dimensionalidad

Clase 3 - 08/05/2023. Practica de reducción de dimensionalidad

Clase 4 - 12/05/2023. Teoría de análisis estadístico de formas

Presentaciones:

- Grupos de 2 o 3
- Máximo 10 minutos
- Temática
 - Aplicación a tema de trabajo real
 - Análisis y presentación de paper → elegir un paper y presentarlo

CIENCIA DE DATOS

- Campo multidisciplinario → abarca muchas áreas. Es extraer información a través de datos.
- Estadística, métodos computacionales e información "área dependiente"
- Objetivo: extraer información de los datos para
 - comprender fenómenos → investigación básica
 - resolver problemas → logística, financieros, ...
 - generar predicciones
 - tomar decisiones (en base a esas predicciones)

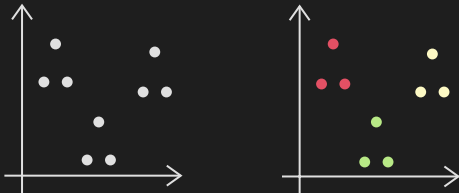
Surge paulatinamente con el desarrollo tecnológico. La cantidad de datos fue creciendo exponencialmente ⇒ extraer información es cada vez más complejo.

APRENDIZAJE AUTOMÁTICO

- extracción de información a partir de un conjunto de datos
- **SUPERVISADO**
 - se trabaja con datos etiquetados
 - * la etiqueta puede ser cualquier cosa que se nos ocurra
 - * reconocimiento de caracteres
 - dataset etiquetado por un método fiable
 - * se usa la palabra supervisado porque necesito tener un conjunto de datos con etiquetas conocidas, creadas con un método confiable. En medicina es muy complicado, porque no hay disponibles muchos datos etiquetados. Es una tarea manual y que lleva mucho tiempo \Rightarrow puede diferir entre un observador y otro. Requiere mucho entrenamiento del observador
 - Entonces, el objetivo es buscar el etiquetado automático.
 - * Aprender cuales son las etiquetas para luego asignarle etiquetas.
 - * Ejemplo: redes neuronales. Se entrena a la red, y luego se evalúa qué tan bien realiza las predicciones el sistema.
 - problemas típicos: clasificación (salida categórica) y regresión (salida continua)
- **NO SUPERVISADO**
 - los datos no poseen etiqueta asociada
 - se busca identificar patrones que permitan separar y organizar los datos. Sin que haya un conocimiento a priori. No se sabe nada de los datos y se busca obtener información sobre ellos.
 - análisis exploratorios
 - problema típico: clustering

CLUSTERING

- agrupación de datos en clusters
- los elementos de un cluster son "similares"
- los elementos de distintos clusters son "diferentes"



- **Aplicaciones:**
 - **Análisis exploratorios**
 - * identificaciones de patrones
 - * compresión de información
 - * generación de modelos
 - **Procesamiento de imágenes**
 - * segmentación de estructuras (Identificar unívocamente pixel por pixel. Si cada pixel corresponde a un objeto de interés. Es de lo más difícil de PDI)
 - * reconocimiento de objetos
 - * tracking (también aplicado a videos. Tracking es para videos)
 - **Marketing**
 - * segmentación de clientes
 - **Detección de anomalías**
 - * patrones inusuales
 - * outliers
 - * errores de producciones

- * fraudes (en caso que se muevan datos económicos)
- * patologías
- **NLP (Natural Language Processing)**
 - * tipos de documentos o discursos
 - * modelos de escritura
 - * registro de información
 - * es posible detectar el tono con el que está escrito
 - * grammarly

El proceso de Cluster me agrupa los datos, pero no me dice nada de esos grupos. Los datos son similares entre ellos pero distintos de estos. Luego está en la persona que analiza qué hacer con esos datos. La interpretación tiene que ser dada por la persona que realiza el análisis

Es importante conocer el problema que se quiere resolver.

TIPOS DE ALGORITMOS DE CLUSTERING

- **Basados en centroides** → el que más se use
 - medias de los clusters
 - divisiones entre clusters lineales y convexas
 - el punto medio de los elementos que perteneces a un cierto cluster.
 - siempre es fácil analizar los datos en 2D, aunque pocas veces es el caso.
 - me permite separar los grupos de datos que estén más distantes unos de otros.
 - diagrama de voronoi → las lineas quedan definidas en base a la posición de los centroides
- **Basados en densidad**
 - formas de clusters arbitrarias (no lineales y convexas)
 - los dos centroides "están superpuestos" ⇒ no se logra separar los dos clusters
 - mira la densidad de los datos
 - que tienen de distinto unos de otros? que en el medio hay sectores de muy baja densidad
- **Jerárquicos**
 - distintos niveles de clusters
 - genera clusters con distinta jerarquía
 - ejemplo: cluster de animales. Jerarquía más baja: berro de raza pug, perro de raza labrador... Jerarquía un poco más alta: perros (solamente)... Jerarquía más alta: animales de 4 patas... y así. Un objeto pertenece a más de un cluster, pero con distintas jerarquías.

Datos categóricos: no es continuo (como la edad), es por ejemplo el sexo biológico

Importante:

- El aprendizaje supervisado tiene etiquetas como el resultado del proceso
- El aprendizaje no supervisado tiene variables categóricas, y son parte de lo que define al propio dato

Qué pasa si cae sobre una linea? → Como todo algoritmo iterativo tengo que tener una condición de finalización

- un dato
- una cantidad de iteraciones

Requieren una **definición de distancia** entre puntos. Siempre vamos a necesitar definir una distancia.

La distancia euclidea es la más difundida

ESPACIO MÉTRICO

conjunto de elementos más una definición de alguna distancia. Todos los espacios \mathbb{R}^n de Álgebra son espacios métricos.

DISTANCIA: cualquier función que cumpla con:

- $d(x, y) \geq 0$
- $d(x, y) = 0 \Leftrightarrow x = y$
- $d(x, y) = d(y, x)$
- $d(x, z) \leq d(x, y) + d(y, z)$ (desigualdad triangular)

Siempre que se cumpla con todo esto, es una distancia.

Ejemplos:

- distancia euclídea: $d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$
- distancia Manhattan: $d(x, y) = |x_1 - y_1| + \dots + |x_n - y_n|$
 - taxicab distance
 - rappi
- distancia geodésica: mínima distancia sobre una superficie
 - no tiene una definición matemática
 - si es un plano o hiperplano, la distancia geodésica coincide con la euclídea
 - aplica a cualquier superficie
- distancia de Chebyshev: $d(x, y) = \max(|x_i - y_i|)$
 - me quedo con el máximo valor absoluto
 - también está pensada para distancias cuadradas
 - pensar en el rey de ajedrez ♔

ALGORITMO K-MEANS

- tipo más simple de clustering basado en centroides
- clasifica a los puntos en k clusters
- cada cluster está representado por un vector n-dimensional (que corresponde al cada centroide)
- cada punto pertenece al cluster más cercano (fundamental la definición de distancia)
- algoritmo iterativo

En este algoritmo la pertenencia es al 100%

Pasos

1. Definición de la cantidad de clusters, k
2. inicialización de los centroides
3. asignación de un cluster a cada punto → surge la definición de distancia
4. recálculo de los centroides
5. repetición de 3 y 4 hasta convergencia

MÉTODOS DE INICIALIZACIÓN

- **Método de Forgy**
 - elección de k puntos aleatorios como centroides
- **Método de partición aleatoria**
 - asignación de cluster aleatorio a cada punto
 - calculo inicial de los centroides como la media de cada cluster
- **Otra opción es asignar puntos aleatorios como centroides**
 - es conveniente hacerlo dentro de una *boundig box* de los datos
 - \approx a Forgy, pero ahora elijo puntos del espacio donde están contenidos los datos del dataset.
 - No necesariamente el centroide sea parte del dataset

VENTAJAS DE K-MEANS

- simplicidad
- escalabilidad
- eficiencia

LIMITACIONES DE K-MEANS

- sensibilidad a las condiciones iniciales \Rightarrow quizás los métodos de inicialización produzcan resultados \neq . Hay gente que estudio, y es todo un tema, ah.
- no permite identificar bordes no lineales
- no permite identificar clusters no convexos (no se puede ir de un punto a otro sin salir del polígono)

\Rightarrow el resultado de los clusters siempre van a ser **polígonos**

CLUSTERING DIFUSO (FUZZY CLUSTERING)

- también llamado soft clustering, en contraste al hard clustering
- asigna una probabilidad de pertenencia o grado de pertenencia a cada uno de los clusters
- generalización del concepto de clustering
- cada punto puede pertenecer a más de un punto simultáneamente
- la pertenencia a cada cluster es un valor entre 0 y 1
- las pertenencias de un punto a todos los clusters suman 1
 - interpretación probabilística
- un hiperparámetro regula cuan difusa es la pertenencia. Cuan difuso es el límite.
 - hiperparámetro porque es lo que está "por arriba"

"El algoritmo no se la juega tanto"

Pasos

- muy similar al k-means
- cambia el modo de cálculo de los centroides

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$$

$$w_k = \frac{1}{\sum_{j=1}^{N_c} (\|x - c_k\| / \|x - c_j\|)^{2/(m-1)}}, \text{ } j \text{ va desde 1 hasta la cantidad de clusters}$$

- k-means: sería lo mismo recalculando los centroides que para el método k-means normal, pero con todos los pesos siendo 0 o 1. Aquí los pesos están entre (0, 1)
- los centroides son promedios de los puntos ponderados por las pertenencias
- $m \in (0, +\infty)$: hiperparámetro que regula. A mayor m , más difusa la clasificación
- En el límite $m \rightarrow 1$, FCM equivale a k-means

ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

- Es un problema de **REDUCCIÓN DE DIMENSIONALIDAD**

EJEMPLO 1 pacientes con datos clínicos, y el dato de presión arterial

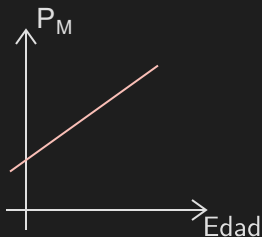
#	P_S	P_D	P_P	$P_M = \frac{1}{2}P_D + P_P$
1				
2				
\vdots				
N				

La P_P y la P_M son combinaciones lineales

- Manipular datos es cada vez más difícil cuanto más grande sea la dimensión
- Cómo hago para identificar cuales son los datos importantes y cuales no? → a eso viene el análisis de componentes principales
- Estoy perdiendo en todos lados:
 - pierdo almacenamiento
 - en una red neuronal, se manipulan datos de dimensión más alta, innecesariamente

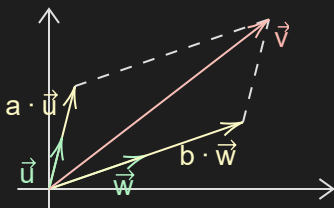
EJEMPLO 2

- edad y P_M también están correlacionadas, pero no es tan fácil notarlo...
- \uparrow Edad $\Rightarrow \uparrow P_M$, pero no es correcto reducir la dimensión en este caso, porque hay dispersiones.
- Qué pasa si yo solo quiero quedarme con **una y solo una** dimensión? "No me quedaría con ninguna, y con las dos"
 \Rightarrow no me quedo ni con la edad, ni con la presión, sino con una combinación lineal de ambas



COMBINACIÓN LINEAL

- $\vec{v} = a \cdot \vec{u} + b \cdot \vec{w}$
- me genera nuevos vectores pero siempre en el mismo espacio originado por los vectores originales



TRANSFORMACIONES LINEALES

- se le dice transformación para indicar el efecto que tiene sobre el espacio
- $\vec{w} = T(\vec{u})$
 - $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$
 - $T(\lambda \cdot \vec{u}) = \lambda \cdot T(\vec{u})$
- Ejemplo: $T(\vec{u}) = \begin{pmatrix} 3u_1 - 2u_2 \\ \frac{3}{2}u_1 + \frac{2}{5}u_2 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ \frac{3}{2} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$
 - Lo que estoy haciendo es un cambio de base!
 - La transformación lineal cambia la base que define el plano
 - $T(\hat{i}) = \begin{pmatrix} 3 & -2 \\ \frac{3}{2} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ \frac{3}{2} \end{pmatrix}$
- Otro ejemplo: $A = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix}$, $\vec{v} = (3, 2)$
 - Que la transformación sea lineal implica que el origen de coordenadas se conserva

- si lo imagino como un conjunto infinito de puntos, el espacio se deforma manteniendo algunas cosas constantes:
 - * las cosas paralelas siguen siendo paralelas
 - * las distancias/ proporciones entre los puntos también se mantiene
- Entonces, $\vec{v} = (3, 2)$ va a ir a parar al: $T(\vec{v}) = A \cdot \vec{v} = (4, 4)$
 - * El $(4, 4)$ sigue estando expresado en la base canónica
 - * $T(\vec{v}) = 4\hat{i} + 4\hat{j} = 3\begin{pmatrix} 2 \\ 0 \end{pmatrix} + 2\begin{pmatrix} -1 \\ 2 \end{pmatrix}$
 - * cuando hago un cambio de base, el vector en sí no cambia sus coordenadas, va a cambiar cuando quiero expresarlo en la forma canónica

• Se dicen lineales porque le hizo una transformación geométrica. Cuanto lo giro y escaló sale del determinante de la matriz A , $\det(A)$

• **DETERMINANTES** Qué es lo que se escala por 4? el área generada por los vectores base del espacio \hat{i} y \hat{j} , al nuevo paralelogramo. El determinante me indica cuánto cambian las áreas de objetos al aplicarle la transformación

• **MATRICES DE ROTACIÓN** Permutar las componentes del vector!

$$- A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} u_2 \\ u_1 \end{pmatrix}$$

– cualquier matriz que cumple con las características de rotar al ángulo

$$- \text{Matriz genérica: } R = \underbrace{\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}}_{\text{norma 1}} \underbrace{\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}}_{\text{norma 1}}$$

– que el determinante sea 1

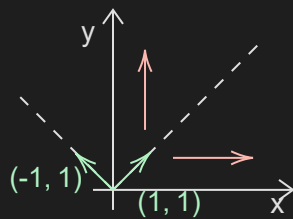
– que sus columnas tengan norma 1, y que sean ortogonales \Rightarrow ser ortonormales

– Acá el vector $(1, 1)$ o cualquier vector que esté sobre la recta del "eje" no es modificado. Mantengo el mismo vector.

* Ejemplo: el $(1, 1) \rightarrow A\vec{v} = 1\vec{v}$

* cualquier vector que se encuentre sobre la otra recta también funciona como autovector

$$\cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (-1) \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



* el que me dice cuánto se esta escalando el autovector es el autovalor asociado

• **DIAGONALIZACIÓN**

– Cuando los autovectores son ortogonales, la inversa coincide con la transpuesta

$$- A = \begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix} = V \cdot \Sigma \cdot V^{-1} \rightarrow V = (\text{autovalor 1} \quad \text{autovalor 2})$$

– Propiedades:

* $A^n = V \cdot \Sigma^n \cdot V^{-1} \rightarrow$ me sirve para ahorrar tiempo de cómputo

* Σ : escala

* V y V^{-1} : rotan

El análisis de componentes principales es descomponer la matriz en sus diagonales!

EJEMPLO CORREDORES

- Datos de 12 corredores de una carrera de 16 km

- En cada una de las columnas están los tiempos en minutos de lo que tardó cada uno en hacer cada tramo

Si quiero describir el **RENDIMIENTO** de cada uno, tenemos varias formas.

Extraer información del espacio de \mathbb{R}^4 es complicado, entonces veamos si podemos hacer algo:

- w_1 es el promedio de la 1ra y 2da \rightarrow 1er tramo de 8km
- w_2 es el promedio de la 3ra y 4ta \rightarrow 2do tramo de 8km
- w_3 es la resta de $w_1 - w_2$

Estoy perdiendo información. El tema es cómo puedo hacer para encontrar las combinaciones lineales de manera tal que se pierda la menor cantidad de información.

Los datos de w_1 y w_2 tiene que ver con "cuanto tardó". w_3 me da información de cómo es el estilo de correr del corredor: si se gasta al principio o al final.

ANÁLISIS DE COMPONENTES PRINCIPALES

- no saca cantidad de dimensiones!
- reexpresa a las variables originales en nuevas variables, donde cada una de esas variables es una combinación lineal de las originales

COMPONENTES PRINCIPALES

- Cargas de la 1) componente principal: cuánto tardó en total, pero normalizado.

$$\sqrt{(0,5)^2 + (0,5)^2 + (0,5)^2 + (0,5)^2} = 1$$

- Cargas de la 2) componente principal: w_3 normalizado
- Cargas de la 3) componente principal
- Cargas de la 4) componente principal

Retomamos el ejemplo de Edad y P_M : "cual de las dos tiene más información"

La combinación lineal ideal va a ser la que esté sobre esa recta

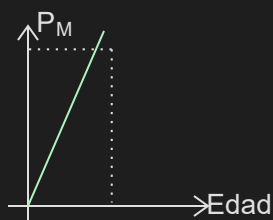


Pierdo menos información
si proyecyo sobre P_M

INFORMACIÓN es poder diferenciar uno de otro. La variable que me da más información es la que abarca más variabilidad (desvío estándar) posible!

Lo que perdí es el rango de información por el cual no proyecto.

Variabilidad o varianza es sinónimo de información



Acá lo que pierdo es cuánto se alejan perpendicularmente de ese vector

Tengo infinitas direcciones perpendiculares.Cuál de todas esas me da más info? → la que tiene mayor varianza posible.
Siempre cada componente principal tiene que ser ortogonal a todas las anteriores y representar la mayor cantidad de varianza posible.



Transformada de ____ es lo mismo que análisis de componentes principales
Es una medida de la correlación lineal de los componentes principales

La covarianza describe el grado de correlación lineal! Entonces:

Covarianza nula \Rightarrow ninguna relación lineal entre las variables
 ninguna relación entre las variables \Rightarrow covarianza nula

MATRIZ DE COVARIANZA

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}, \sigma_x^2 = E[(x - E(x))^2], \sigma_{xy} = E[(x - E(x))(y - E(y))]$$

Una diferencia al cuadrado me da la noción de distancia.

La esperanza es un concepto teórico que describe la población, mientras que el valor medio es un estimador de la esperanza.
Esperanza: es el valor esperado.

Covarianza: valor medio de las diferencias al cuadrado, entre cada valor y la media. Es cuánto se alejan los datos de la media, en promedio

PLANTEOS

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ x_{21} & & x_{2m} \\ x_{31} & & x_{3m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \Rightarrow \bar{X} = [\bar{x}_1 \quad \cdots \quad \bar{x}_m] \xRightarrow{\text{creo}} X_m = \begin{bmatrix} \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_1 & \cdots & \bar{x}_m \end{bmatrix} \text{ (n filas)}$$

Ahora, lo que hago es $X_Z = X - X_m$

Es la matriz de datos, toda movida para que el valor medio de los datos coincida con el origen.

Se hacen más fáciles los cálculos de la matriz de covarianza.

Si ahora hago: $X_Z^T \cdot X_Z = \Sigma$, matriz de covarianza sigma

- cuando coinciden fila y columna (diag principal): es la varianza, $\sigma_x^2 = E[(x - E(x))^2]$

$$- \sigma_x^2 = \sum_{i=1}^N (x_{i1} - \bar{x}_1)^2$$

- cuando no coinciden: es la covarianza, $\sigma_{xy} = E[(x - E(x))(y - E(y))]$

$$- \sigma_x^2 = \sum_{i=1}^N (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$$

Las componentes principales son los autovectores de la matriz de covarianza!

Siempre voy a tener m autovectores linealmente independientes. Porque es definida positiva, entre otras cosas

El valor de cada autovalor está directamente relacionado con la varianza que se obtiene al proyectar los datos sobre el autovector asociado

Los autovectores de 3 y 4 me están aportando menos del 5% de la **variabilidad** total de los datos \Rightarrow los descarto

Si yo tomo los 4 autovectores, llego al 100% de la variabilidad

Yo puedo descartar algunas de las dimensiones, pero conservar gran variabilidad.

Mis cuatro variables originales me daban cuanto tardó en cada tramo. Me da información, pero escondida. solo con los datos:

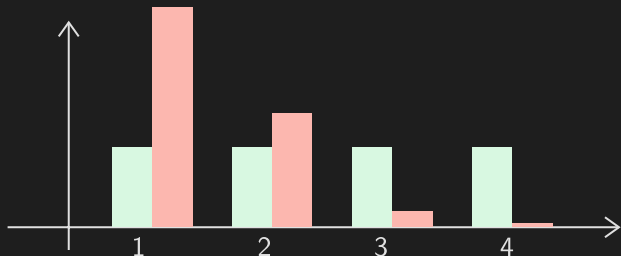
- rapidez global
- estilo de correr

ya puedo distinguir uno del otro, perdiendo menos del 5% de lo que los diferenciaba originalmente

TEST DE ESFERICIDAD

- una esfera "abstracta"
- la dirección de máxima variación \rightarrow no tiene una, si es una esfera no hay una única dirección más importante que la primera
- si los datos no se distribuyen esféricamente, tengo que seguir evaluando.
- Al principio, es todo no esférico. Pero al final, es más o menos parejo en todas las direcciones \Rightarrow "ahí cortar"

Cual es el cambio de base ideal, optimo, para poder transformar los datos a una nueva base de manera tal que cada una de las variables conserve la mayor cantidad de información



Teníamos

- 1 componente: de escala. Suele asociarse con componentes de escala.
- 2 componente: de forma. Suele asociarse con componentes de forma.

EJEMPLO: AORTA

- Tamaño
- Arco de la aorta
- "hacia donde apunta el arco"

Center line \rightarrow línea media de la aorta, se ve cómo se agranda o se achica, cómo se modifica

Los autovalores son variabilidad

DESCOMPOSICIÓN EN VALORES SINGULARES (SVD)

- Generalización de PCA

EJEMPLO

Una imagen es una matriz de datos. Por cada valor de la matriz tengo la información del valor de gris.
Una matriz es una grilla de pixeles.

Matlab ya tiene la función `PCA()` y `svd()`

$[U \ S \ V] = \text{svd}(\text{IMG})$, IMG es la matriz de pixeles

- S matriz diagonal, rectangular del mismo tamaño que la imagen
 - es análogo de la descomposición de autovectores y autovalores (están ordenados en magnitud)
 - S: "valores singulares" → generalización de autovalores (la palabra autovector y autovalor se usa solo si la matriz es cuadrada)
- U: 250x250
 - matriz de vectores singulares "a la izquierda"
- V: 256x256
 - matriz de vectores singulares "a la derecha"
- Entonces: a la matriz IMG la voy a descomponer. $\text{IMG} = S \cdot U \cdot V$

Me interesa buscar cuál es el primer valor principal que sea más chico que 1 → los menores a 1 los descarto! (recordar de PCA)

A ese valor lo guardo en la variable `idx`

Creo una matriz $S_2 = S$, solo que a partir del elemento 71 es toda cero.

La imagen se conserva, usando mucha menos información que la imagen original.

Reducir la dimensionalidad (no en el mismo sentido que PCA), pero sí reducir la cantidad de información que necesito. La ventaja de hacer PCA es que "selecciono" la información que pierdo. Pierdo menos información.

Manera de comprimir una imagen (el más conocido es el JPG)

MATRIZ DE CORRELACIÓN

- En el caso que tomara la misma variable en x e y (autocorrelación) → siempre tiene 1 en la diagonal
- fuera de la diagonal es el R (que va entre -1 y 1)
- Qué diferencia hay entre hacer una matriz de covarianza y otra usando la de correlación. Son las mismas dimensiones de las matrices.
- si uso esta se sesga el resultado por las escalas. A todas las tengo en las mismas escalas
- cuando las escalas son muy diferentes unas con las otras uso la matriz de correlación para que no pase que una con una escala muy grande

ANÁLISIS ESTADÍSTICO DE DATOS (SSF)

TRANSFORMACIONES SEMEJANTES

- Lineales:
 - escalado: $\hat{X} = s \cdot \bar{X}$, es una matriz escalar, con todos los elementos iguales. Escalado uniforme, se escala a todas las dimensiones por igual
 - rotación, $\hat{X} = R \cdot \bar{X}$ es una matriz rotacional, con R ortonormal. Columnas tienen norma 1, y forman una base ortogonal
- No lineales
 - traslación: $\hat{X} = \bar{X} + \bar{T}$, no es lineal porque se pierde el origen

El vector \hat{X} es \approx al vector X . Se conserva la proporcionalidad. Conservo la relación entre las distancias de las distintas dimensiones.

Por qué a veces pueden ser cuatro?

- Reflexión
 - espejarlo
 - implícito en el factor de escala negativo. A veces se cuenta la reflexión como algo distinto

Son aquellas transformaciones que conservan la forma de un objeto

QUÉ DEFINE UNA FORMA

- Contorno
 - Cuándo deja de ser una forma? \rightarrow un objeto es una cierta forma y cuando un objeto es otra forma
 - Información geométrica de un objeto que permanece luego de eliminar:
 - rotación
 - escala
 - posición
- \Rightarrow **La forma es invariante ante transformaciones semejantes**
- Descripción de una forma: similitud con objetos conocidos
 - No utilizable en modo algorítmico

LANDMARKS

Los que se suele utilizar es la definición de landmarks

Permiten definir matemáticamente una forma

Puntos de un objeto que se corresponden con un punto en otro objeto del mismo espacio.

Punto de correspondencia que coinciden en cada sujeción de una población. Correspondencia semántica en el significado que tiene ese punto.

Tipos:

- anatómicos
- matemáticos: rescata alguna propiedad matemática de una curva o cualquier gráfica de una función n -dimensional. Cualquier característica matemática que se puede definir en una gráfica.
- pseudolandmarks: un término que se inventó para meter adentro todo lo que queda.

FORMA: cada vector de landmarks

Matemáticamente, forma definida por n puntos en el espacio \mathbb{R}^n es un vector nk -dimensional.

Ejemplo: $k = 2$: $\bar{X} = [x_1 \dots x_n ; y_1 \dots y_n]^T$

Los landmarks son puntos que viven en el espacio que vive la forma, y el conjunto de los landmarks son los que van a definir una forma.

Objeto: es más genérico, incluye muchas formas.

CÓMO SE IDENTIFICA UN OBJETO

- En el espacio de las formas un decágono regular si lo represento con 10 landmarks es la misma forma que una circunferencia.
1. Superponerlos
 - (a) escalando \rightarrow qué es el tamaño

(b) desplazar a la misma posición → qué es la posición

(c) rotarlos → qué es la inclinación

2. ANÁLISIS DE PROCRUSTES "el estirador"

ESPACIO DE FORMAS

- Matemáticamente: conjunto de n puntos de \mathbb{R}^k no idénticos, resultantes de transformaciones semejantes
- todos los conjuntos de las posibles formas de un objeto
+ definición de distancia = espacio métrico
- tengo que definir una distancia entre formas → para poner un umbral y saber cuándo una mano deja de ser una mano

TAMAÑO

POSICIÓN

ROTARLOS

A esta altura ya eliminamos factor de escala, posición y rotación

DISTANCIA ENTRE FORMAS

- **Distancia de Hausdorff:**
 - calcular la distancia entre subconjuntos, que pueden tener infinitos elementos.
 - necesariamente tengo que tener una definición de distancia entre los puntos.
 - tiene una def matemática concreta
 - la mayor distancia que separa un punto de uno de los subconjuntos con el punto más cercano del otro subconjunto
- **Distancia de procrustes**
 - más específica
 - la cantidad de elementos en cada subespacio tiene que ser la misma
 - tiene que haber una definición de distancia para cada uno de los elementos
 - es básicamente como una norma: $P_d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + \dots}$
 - es la distancia euclídea generalizada al espacio n -dimensional donde viven estas formas
 - ¿cómo se realiza cada paso del alineamiento?
 - * calcular el centroide, centro de masa.
 - * $(\bar{x}, \bar{y}) = \left(\frac{1}{n} \sum_i x_i; \frac{1}{n} \sum_i y_i; \dots \right)$
 - * no vive en el espacio de las formas, sino en el espacio de los landmarks

Cálculo de la escala

- tamaño de una forma
- cualquier valor positivo que cumple con la siguiente propiedad:
 - esta mano es dos veces más grande que la otra \Rightarrow cuando calculo el tamaño me tiene que dar 2 veces.
 -
- Se suele usar la norma 2:
 -

Cálculo de la rotación

- cómo se alinean las formas: se suele hacer un PCA
- multiplico la matriz de forma1 y forma2, transponiendo una de las dos, para que me quede una matriz de $k \times k$

- La rotación óptima se puede calcular como: VU^T

Si tengo n formas, se hace un proceso iterativo donde se agarra una forma cualquiera (forma media/ forma con orientacion 0) y se alinea cada una de las formas a aesa, luego se calcula la forma nueva denuevo. Si cambió la estimacion de forma media, repetir.

Qué es la forma media? la media de procrustes $\bar{X} = \frac{1}{N} \sum_{i=1}^N \bar{X}_i$, N es la cantidad de formas que tenemos

Qué sucede con la normalización (proceso de alineamiento: escala, rotacion, traslación)?

Proyeccion al espacio tangente

[imagen]

para que la distancia no este dada con la forma geodésica.

ejemplo

los puntitos son las variaciones de los landmarks de todas las formas

la proyeccion al espacio tangente lo que hace es linealizar el espacio, para no tener que calcular las distancias sobre una esfera.

