Cuando termine el curso, sería importante que sepan tener respuestas a las siguientes preguntas sobre los distintos temas abordados en el curso. Estas preguntas tienen dos objetivos.

1- Ayudar a redondear conceptos de cada uno de los temas.

2- Preparación de la parte de preguntas teóricas el día del examen.

Estas preguntas no forman parte del material que deben entregar. Lo que deben entregar son lo que les indico en el otro topic.

Iré actualizando esta lista en la medida que recorramos los próximos temas.

**ÁLGEBRA LINEAL**

1.     **¿Qué es una forma cuadrática?**

**2.     ¿Qué propiedades tienen los valores propios de una matriz simétrica?** Los valores propios de una matriz simétrica son todos reales, y los vectores propios correspondientes son ortogonales entre sí.

**3.     ¿Qué significa que una matriz simétrica sea definida positiva o semidefinida positiva? Indicar dos definiciones equivalentes entre sí.**

**4.     ¿Qué relación existe entre la traza de una matriz y sus valores propios? ¿Y entre el determinante de la matriz y sus valores propios?** La traza de una matriz es igual a la suma de sus valores propios. El determinante de una matriz es el producto de sus valores propios.

**5.     ¿Qué dice el teorema espectral? Dar dos resultados equivalentes entre sí.** El teorema espectral establece que cualquier matriz simétrica real puede diagonalizarse mediante una matriz ortogonal. Es decir, A=QΛQ^T , donde Q techo es una matriz ortogonal cuyas columnas son los vectores propios y Λ techo es una matriz diagonal con los valores propios.

**6.     ¿Qué dice el teorema de la esfera unidad?**

**7.     ¿Qué propiedades tienen las matrices de covarianzas?** Las matrices de covarianzas son simétricas y semidefinidas positivas, lo que significa que todos sus valores propios son no negativos. Además, la diagonal contiene las varianzas de las variables.

**ANÁLISIS EN COMPONENTES PRINCIPALES (PCA)**

1**.     ¿Cuál es el objetivo principal que aborda el análisis en componentes principales?** El objetivo es reducir la dimensionalidad de los datos, manteniendo la mayor cantidad posible de la variabilidad total. Esto se logra transformando las variables originales en un conjunto más pequeño de variables no correlacionadas llamadas **componentes principales**.

2.     **¿Cómo se hallan las componentes principales? ¿Qué problema resuelven?** Las componentes principales se obtienen calculando los valores y vectores propios de la matriz de covarianzas o la matriz de correlaciones de los datos. Resuelven el problema de maximizar la varianza explicada por las nuevas variables (componentes) con respecto a las originales.

**3.     ¿Qué problema trae aparejada la existencia de datos atípicos en el PCA?** Los datos atípicos pueden distorsionar los resultados del PCA, ya que las componentes principales están basadas en la varianza, y los atípicos pueden inflar la varianza de manera desproporcionada.

**4.     Si las variables tienen distintas unidades de medida o distintas magnitudes ¿qué problema puede aparecer en el PCA y cómo se puede hacer para resolverlo?** Si las variables tienen diferentes escalas, las que tienen mayores magnitudes dominarán las componentes principales. Para resolver este problema, se estandarizan las variables, es decir, se les da la misma escala convirtiendo sus medias en 0 y sus desviaciones estándar en 1.

**5.     ¿Qué desventaja puede tener estandarizar los datos antes de hacer un PCA?** Al estandarizar, se pierde la interpretación directa de la importancia de las variables originales, ya que todas las variables contribuyen por igual a la varianza total, independientemente de su escala original.

**6.     ¿Qué es un biplot y para qué sirve?** Un biplot es una representación gráfica que muestra simultáneamente tanto las observaciones como las variables en el espacio definido por las componentes principales. Permite interpretar las relaciones entre las observaciones y cómo las variables influyen en dichas observaciones.

**7.     ¿Cómo se interpretan en el biplot el lugar que ocupan las observaciones? ¿Y el ángulo entre los vectores? ¿Y             la longitud de los mismos?**

* El lugar de las observaciones en el biplot refleja sus relaciones con las componentes principales.
* El ángulo entre los vectores de las variables indica la correlación entre ellas: ángulos pequeños significan alta correlación positiva, ángulos rectos indican no correlación y ángulos grandes indican correlación negativa.
* La longitud de los vectores refleja la importancia de las variables en la varianza explicada por las componentes.

**8.     ¿Cómo se sabe cuándo un biplot es mucho o poco  informativo del conjunto de datos que tenemos?** Un biplot es más informativo cuando las primeras dos componentes principales explican un alto porcentaje de la varianza total de los datos. Si explican poco, el biplot será menos representativo del conjunto de datos completo.

**9.     Cuando vemos la salida de la  función prcom() en R ¿qué propiedades cumplen los valores que nos arroja la matriz de rotación**? Los valores en la matriz de rotación son los coeficientes que definen cómo las variables originales se combinan para formar las componentes principales. Esta matriz es ortogonal, lo que significa que sus columnas son ortonormales (vectores propios).

En Python, el equivalente a la función prcomp() de R, que se utiliza para realizar Análisis en Componentes Principales (PCA), es la clase PCA de la biblioteca scikit-learn. Esta clase proporciona una funcionalidad similar para realizar PCA y generar la matriz de rotación.

*# Obtener la matriz de componentes principales (rotación) rotacion = pca.components\_*

***¿Qué es la matriz de rotación (o components\_ en Python)?***

*La* ***matriz de rotación*** *(que en scikit-learn se llama components\_) es una matriz cuyas filas son los* ***vectores propios*** *de la matriz de covarianzas o correlaciones de los datos originales. Estos vectores propios definen las direcciones de las nuevas componentes principales.*

* *En el contexto de Python,* ***pca.components\_*** *es equivalente a la* ***matriz de rotación*** *que devuelve prcomp() en R.*
* *Las columnas de esta matriz representan cómo las* ***variables originales*** *se combinan linealmente para formar las nuevas componentes principales.*
* *Al igual que en R, esta matriz es* ***ortonormal****, es decir, sus columnas son vectores propios y son ortogonales entre sí.*

***Interpretación de la salida***

* ***pca.components\_****: Cada fila de esta matriz corresponde a un vector propio asociado a una de las componentes principales. Los coeficientes en estas filas nos indican las ponderaciones (o combinaciones lineales) de las variables originales que forman cada componente principal.*
* ***Ortonormalidad****: Igual que en prcomp(), las filas de la matriz en Python son ortogonales, lo que significa que las componentes principales son no correlacionadas entre sí. Además, las longitudes de estos vectores son 1 (norma 1), garantizando la ortonormalidad.*

***Propiedades importantes:***

1. ***Ortogonalidad****: Las componentes son ortogonales entre sí, lo que asegura que cada nueva componente captura una parte independiente de la variabilidad de los datos.*
2. ***Varianza explicada****: Se puede acceder a la varianza explicada por cada componente principal usando pca.explained\_variance\_ratio\_. (Esta salida indica qué porcentaje de la varianza total es capturada por cada componente principal, lo que también es una parte clave del análisis.)*