

ITAM  
Métodos Numéricos y Optimización  
(MAT-34420)

Práctica 2  
Álgebra Lineal Numérica

J. Ezequiel Soto S.  
[ezequiel.soto@itam.mx]

**Parte I. SVD & PCA**

Esta actividad es obligatoria.

*Análisis de Componentes Principales (PCA)*

Elige una de las siguientes bases de datos:

Breast Cancer Wisconsin (Diagnostic) Data Set:

<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/>

Red Wine Quality

<https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>

Human Activity Recognition with Smartphones

<https://www.kaggle.com/datasets/uciml/human-activity-recognition-with-smartphones>

1. Prepara los datos: para cada variable, centra alrededor del cero restando las medias y estandariza la escala dividiendo entre la varianza ( $\text{np.std}$ ). Haz un mapa de calor con la matriz de correlaciones de la base de datos estandarizada.
2. Realiza la reducción de dimensionalidad mediante PCA calculando los eigenvalores y eigenvectores de la matriz de varianza-covarianza ( $A^T A$ ).
3. Calcula la varianza total acumulada en las primeras  $k$  componentes principales.
4. Analizando la forma de calcular la proyección ( $Y = AV$ ), identifica los coeficientes de cada variable en las primeras componentes principales. Discute lo observado: ¿cuáles son las variables que más contribuyen con cada componente? Compara estos resultados con la matriz de correlaciones del paso 1.
5. Grafica los datos en el(los) plano(s) de las primeras componentes principales, las que hayas elegido de acuerdo al criterio del paso 3. Colorea los puntos con sus categorías correspondientes: diagnóstico, calidad o tipo de actividad, respectivamente. Discute: ¿la reducción de dimensionalidad parece facilitar tareas de clasificación de los datos?
6. Grafica las variables en el *espacio latente*: el de sus coeficientes en cada una de las componentes principales. Interpreta.

## Parte II. Mínimos cuadrados

Lee completamente las propuestas de actividad y elige una de ellas.

### *Superficies de curvatura mínima*

Calcula la superficie  $z = f(x, y)$  de área mínima y curvatura constante en distintas configuraciones de frontera. Utiliza el operador:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

como medida de la curvatura. Considera  $(x, y) \in \Omega = [-1, 1] \times [-1, 1]$ . Representa el dominio a través de una malla rectangular regular y utiliza las fórmulas de diferencias finitas.

*Regresión* Dados  $n$  puntos  $(x_i, y_i)$ , el objetivo es encontrar el polinomio de grado máximo  $p$  que minimice la suma de los errores cuadrados:

$$\min \sum_{i=1}^n (y_i - f(x_i))^2$$

Discute los casos:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_p x^p$$

$$f(x) = a_0 + a_1 \cos(x) + a_2 \cos(2x) + \dots + a_p \cos(px)$$

### *Regresión generalizada*

Dados  $n$  puntos  $(x_i, y_i)$ , el objetivo es encontrar el polinomio de grado máximo  $p$  que minimice la suma de los errores cuadrados:

$$\min \sum_{i=1}^n (y_i - f(x_i))^2$$

Discute los casos:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_p x^p$$

$$f(x) = a_0 + a_1 \cos(x) + a_2 \cos(2x) + \dots + a_p \cos(px)$$

## *Evaluación*

[50%] Organiza tus soluciones de la práctica en un cuaderno de python (\*.ipynb), incluye gráficas. Recuerda la diferencia entre discutir una solución y copiarla, siempre opta por lo primero.

[50%] Comunica tus resultados en un documento usando texto, figuras y tablas. Este es un ejercicio de comunicación distinto del cuaderno de python, cuida la calidad.

Si utilizas literatura (formal o informal), incluye las referencias.

Fecha de entrega: **17 de noviembre de 2024**. Dos archivos: .pdf, .ipynb