



Mini-proyecto 2: Modelos de Variable Latente para el Reconocimiento de Emociones

Tratamiento de Señales III
Facultad de Ingeniería
Universidad de Antioquia
Hernán F. García PhD
2023-2

1. Para el sistema de Análisis Multivariado de Señales fisiológicas obtenido en el Primer Mini-Proyecto, se debe construir un módulo de variable latente utilizando el algoritmo de PCA. La idea es proyectar la matriz característica obtenida $\mathbf{X} \in \mathbb{R}^{N \times D}$, a un espacio latente de menor dimensión $q \ll D$, de tal forma que los nuevos datos en este espacio sean $\mathbf{Z} \in \mathbb{R}^{N \times q}$. Selecciones diferentes configuraciones del número de componentes q y evalúe el mejor desempeño arrojado por el modelo midiendo, el *accuracy* y la matriz de confusión. Para ellos:
 - a) Realice la visualización de las componentes en el espacio latente para diferentes configuraciones de q
 - b) Qué puede observar al respecto?
 - c) Entrene un modelo lineal de clasificación **Mínimos Cuadrados**, **Bayesiano** y **Regresión Logística**, con el fin de evaluar el desempeño en el reconocimiento de los tipos de movimiento tanto en el espacio de entrada (sin utilizar PCA), cómo en el espacio latente.
2. Dado que PCA es un modelo lineal de variable latente en el cual las características son mapeadas de la forma $\mathbf{z} = \mathbf{W}\mathbf{x}$, en muchas aplicaciones reales las relaciones entre los datos son no lineales. Es por esto que modelos cómo la reducción de la dimensionalidad a través del uso de funciones de mapeo (kernel) no lineales puede ayudarnos en la tarea. Aquí el mapeo se realiza utilizando $\mathbf{z} = \Phi\mathbf{x}$, donde Φ es la matriz de mapeo no lineal conocida como función de kernel. Estas funciones de kernel pueden tener diferentes naturaleza como lineales, exponenciales, polinomiales, entre otras. Su objetivo aquí es realizar el procedimiento de mapeo del dataset de movimientos utilizando Kernel PCA (Kernel Principal Component Analysis).
 - a) Cómo lucen las características proyectados en ambos espacios latentes PCA y KPCA?. Implemente su visualización.
 - b) Utilice este nuevo subconjunto de datos con un clasificador basado en regresión Logística para el reconocimiento. Qué puede evidenciar en el desempeño del nuevo modelo de reconocimiento (con KPCA) para esta tarea?

NOTA: Los notebooks junto con sus respectivos análisis deben ser compartidos en un link de Github para el cual debe existir un commit intermedio para observar el avance en el desarrollo. Subir en la plataforma de Classroom dicho link de Git donde estará el repositorio de la actividad. Entregar en el classroom hasta el 16 de noviembre 23:59 hrs

éxitos!!!!