



1. DATOS BÁSICOS DEL TFG:

Título: Análisis de componentes principales: fundamentos teóricos y aplicaciones en ciencia de datos

Descripción general (resumen y metodología):

Este Trabajo Fin de Grado tiene como objetivo introducir al estudiante en el estudio del Análisis de Componentes Principales (PCA), una técnica fundamental en el procesamiento de datos, reducción de dimensionalidad y visualización en aprendizaje automático y estadística multivariante.

Se abordará el desarrollo teórico del PCA desde dos perspectivas equivalentes: la formulación basada en la máxima varianza de la proyección y la formulación basada en el error mínimo de proyección. Posteriormente, se estudiarán aplicaciones prácticas relevantes, como la compresión de datos, el “blanqueo” de datos (whitening) y los retos asociados al tratamiento de datos en espacios de alta dimensión.

Además del estudio teórico, se llevará a cabo una implementación práctica de la técnica utilizando herramientas de análisis numérico, aplicándola a conjuntos de datos concretos. Esta parte experimental permitirá analizar empíricamente la eficacia de PCA en tareas de reducción de dimensionalidad y preprocesamiento.

La metodología combinará el estudio bibliográfico, el estudio matemático del problema y la programación práctica en Python u otros entornos adecuados para la manipulación de datos.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

1. Comprender los fundamentos teóricos del Análisis de Componentes Principales (PCA).
2. Estudiar y comparar las formulaciones de máxima varianza y mínimo error de proyección.
3. Analizar las aplicaciones de PCA en el área de ciencia de datos: compresión y preprocesamiento.
4. Implementar y aplicar PCA sobre conjuntos de datos reales y sintéticos, interpretando los resultados obtenidos.

Bibliografía básica:

Bishop, C. M., & Bishop, H. (2023). Deep learning: Foundations and concepts. Springer Nature.

Bishop, C. M. (2006). Pattern recognition and machine learning. New York: Springer.

Kaggle tutorials for Python programming: <https://www.kaggle.com/learn>

Recomendaciones y orientaciones para el estudiante:

Trabajar en el proyecto a lo largo de todo el curso, con la dedicación horaria establecida por la carga de créditos correspondiente.

Tener reuniones mensuales con el supervisor para comprobar el avance del trabajo.

Proporcionar una versión preliminar del trabajo con suficiente antelación a la fecha de depósito, para poder recibir correcciones por parte del supervisor.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: PABLO MORALES ÁLVAREZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: pablomorales@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: ELENA FERNANDEZ CARO

Correo electrónico: elenafdezcaro@correo.ugr.es