



1. DATOS BÁSICOS DEL TFG:

Título: Análisis de técnicas de explicabilidad para arquitecturas neuronales profundas en clasificación de imágenes

Descripción general (resumen y metodología):

Las redes neuronales son estimadores universales de funciones y han experimentado un gran avance en capacidad en los últimos años debido a su flexibilidad y el incremento de la capacidad computacional disponible, permitiendo usos cada vez más complejos. A medida que las redes neuronales profundas se utilizan cada vez más para la clasificación de imágenes, surge la necesidad de comprender y explicar matemáticamente sus decisiones. Este TFG explora y compara diferentes técnicas de explicabilidad desde una perspectiva matemática, con el objetivo de identificar métodos que puedan proporcionar conocimiento claro y útil sobre el funcionamiento interno de estas arquitecturas complejas que no son directamente entendibles por el ser humano. La metodología de este trabajo se basa en la búsqueda bibliográfica exhaustiva y la investigación experimental. A continuación, se detalla cada uno de estos aspectos:

- **Selección de Fuentes:** Se realizará una búsqueda exhaustiva de literatura científica en bases de datos reconocidas como IEEE Xplore, Springer, Google Scholar y arXiv. Se seleccionarán artículos, conferencias, tesis y libros relevantes que aborden tanto la interpretabilidad de redes neuronales profundas como las técnicas específicas de explicabilidad.
- **Revisión Sistemática:** Se llevará a cabo una revisión sistemática de la literatura para identificar, evaluar y sintetizar los hallazgos más relevantes sobre las técnicas de explicabilidad aplicadas a la clasificación de imágenes. Se prestará especial atención a estudios comparativos y revisiones que analicen múltiples técnicas de explicabilidad.
- **Clasificación y Análisis:** Los estudios encontrados se clasificarán según las técnicas de explicabilidad utilizadas (por ejemplo, Grad-CAM, LIME, SHAP) y se analizarán en términos de su aplicabilidad, efectividad y limitaciones. Se elaborará una tabla comparativa para resumir las características clave de cada técnica, incluyendo métricas de desempeño y áreas de aplicación.

Además, se contemplará la evaluación empírica de varias técnicas de explicabilidad (como Grad-CAM o LIME) sobre arquitecturas convolucionales populares.

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

Objetivos planteados:

- Explorar los fundamentos teóricos de las redes neuronales (en particular, redes convolucionales) y describir sus propiedades matemáticas.
- Estudiar los fundamentos teóricos de diferentes técnicas de explicabilidad.
- Explorar la aplicación de técnicas de explicabilidad en técnicas modernas en machine learning.

Bibliografía básica:

- [1] Christopher M. Bishop. Pattern Recognition and Machine Learning. (Springer, 2006)
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, (MIT Press, 2016). Available at <http://www.deeplearningbook.org>.
- [3] Li, Zewen, et al. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 2021, vol. 33, no 12, p.

6999-7019.

[4] Angelov, Plamen P., et al. "Explainable artificial intelligence: an analytical review." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11.5 (2021): e1424.

Recomendaciones y orientaciones para el estudiante:

Se recomienda al alumno haber cursado las siguientes asignaturas:

Inferencia estadística, Análisis Matemático, Modelos matemáticos II, Métodos Numéricos.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: JULIÁN LUENGO MARTÍN

Ámbito de conocimiento/Departamento: CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

Correo electrónico: julianlm@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:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: Miquel Jiménez Flaquer

Correo electrónico: jimeenez15@correo.ugr.es