Buenas tardes. A continuación, voy a presentar el clasificador de eficiencia que hemos desarrollado, que se trata de un trabajo de Juan Aparicio, José Luis Zofío, Víctor España y yo, Ricardo González.

Diap Intro 1

Como sabéis, el campo del análisis de eficiencia y la evaluación de unidades de toma de decisión ha experimentado avances significativos. Una de las metodologías más destacadas para esto es el Análisis Envolvente de Datos (DEA por sus siglas en inglés), que evalúa la eficiencia considerando múltiples inputs y outputs sin imponer supuestos restrictivos sobre formas funcionales o distribuciones subyacentes.

Sin embargo, los enfoques tradicionales de DEA pueden presentar limitaciones a la hora de capturar patrones y estructuras complejos. Además, existe un riesgo de sobreajuste en el modelo, que puede capturar ruido o características idiosincráticas de los datos en lugar de las verdades relaciones subyacentes, sobre todo cuando existen conjuntos de datos de alta dimensión o cuando el número de DMUs es relativamente bajo en comparación con el número de inputs y outputs.

Además, los modelos tradiciones de DEA dependen de técnicas de programación lineal que pueden no captar adecuadamente las relaciones no lineales y las interacciones input/output. Otra limitación del DEA es su naturaleza determinista, el score de eficiencia no tiene en cuenta las incertidumbres o la variabilidad inherente a los sistemas del mundo real, no reconoce la naturaleza estocástica de muchos procesos de toma de decisiones.

Diap 2 ML

Con el desarrollo de las técnicas de aprendizaje automático, surge la oportunidad de mejorar el DEA aprovechando el poder computacional y flexibilidad que ofrecen, buscando aumentar la precisión, robustez e interpretabilidad de las evaluaciones de eficiencia, buscando captar esas relaciones no lineales que pueda haber.

Se han establecido algunos puentes entre el aprendizaje automático y el DEA. Existen dos corrientes en la literatura:

La primera se centra en adaptar técnicas de ML existentes para garantizar que la función predictiva, una función de producción para nosotros, cumpla con restricciones de forma (monotonía o concavidad). Es decir, adaptan técnicas como SVM, redes neuronales, árboles de clasificación para estimar una función de producción.

La segunda corriente adopta un enfoque en dos etapas. En la primera etapa, se aplica un DEA para obtener los scores de eficiencia de cada observación. En la segunda etapa, los scores de eficiencia se tratan como la variable respuesta en un modelo estándar de aprendizaje automático.

Ambas corrientes han tendido puentes entre el DEA y el aprendizaje automático, pero siguen existiendo brechas. Nuestra investigación aporta una nueva metodología que está relacionada con la segunda corriente mencionada. Los riesgos que existen al utilizar esta segunda corriente, y que pretendemos abordar, es su utilización en contextos inciertos y ruidosos y al utilizar los scores ofrecidos por DEA, la evaluación de la eficiencia no se mejora al incorporar técnicas de aprendizaje automático, la clasificación es la misma.

Diap 3 Types

Por tanto, por primera vez, utilizaremos un modelo de clasificación en vez de regresión en la segunda etapa que combina DEA y aprendizaje automático.

Brevemente, existen dos tipos principales de aprendizaje en el aprendizaje automático. El supervisado, con etiquetas, y el no supervisado, sin etiquetas. Nuestro modelo de clasificación se encuentra en el aprendizaje supervisado y, en función de los datos, determinará las dos regiones y ofrecerá una clasificación y una probabilidad de pertenecer a esa clase, a diferencia del un modelo de regresión que, según los datos, estima el valor de la variable que se desea estimar.

Las técnicas de aprendizaje automático de clasificación que hemos utilizado en nuestra investigación son Support Vector Machine y Redes Neuronales. Primero comentaré en qué consiste SVM:

Diap 4 SVM

Las máquinas de soporte vectorial identifican un hiperplano que separa los datos en dos clases. Este hiperplano se posiciona de manera que maximiza el margen, es decir, la distancia perpendicular entre el hiperplano y los puntos más cercanos de cada clase, conocidos como vectores de soporte.

Una de las características de SVM es la capacidad de utilizar kernels para transformar el espacio de características y elevarlo a uno mayor, donde sí que es posible separar los datos linealmente. Algunas de las más utilizadas es la lineal, la polinomial, la radial o la sigmoide. No obstante, la efectividad de las SVM depende en gran medida de la correcta selección de hiperparámetros, como el parámetro de regularización y el tipo de núcleo. Aquí es donde la validación cruzada juega un papel crucial, ya que permite ajustar estos parámetros para mejorar el rendimiento del modelo y evitar el sobreajuste.

Diap 5 NN

Las redes neuronales han crecido enormemente de popularidad en los últimos años. Estas se caracterizan por imitar el cerebro humano al estar formadas por capas con nodos, llamadas neuronas. Consta de dos procesos, el primero es la propagación hacia delante.