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 kernel. 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 donde los datos se transmiten capa a capa, aplicando un conjunto de pesos y funciones de activación que producen una salida. La función de activación permite capturar relaciones complejas y algunas de ellas son la sigmoidal, la tangente hiperbólica o la ReLU.

En cada época, cada ciclo, se busca minimizar la función de pérdida. Así que, una vez determinado el error, se ajustan los pesos de toda la red para intentar acertar en la predicción.

El rendimiento de las Redes Neuronales depende de la selección de hiperparámetros, como el número de capas, el número de neuronas por capa, la tasa de aprendizaje y los parámetros de regularización. La correcta elección de hiperparámetros busca mejorar el rendimiento del modelo y prevenir problemas como el sobreajuste.

Algunos hiperparámetros buscan determinar la estructura de la red: como el número de capas ocultas y número de neuronas por capa, el porcentaje de neuronas que se desean desactivar entre capas, determinar los pesos iniciales o la función de activación.

Otros hiperparámetros determinan cómo se entrenará la red, qué criterios debe seguir. Como la tasa de aprendizaje, el número de épocas, si se retira un porcentaje diferente de observaciones en cada época y después de actualizar los pesos se vuelven a introducir.

Diap 6 XAI

Ambas técnicas las emplearemos para determinar dos regiones, eficientes y no eficientes, y esto lo conseguiremos mediante un análisis contrafactual que busca: ¿Cómo actuaría el modelo si me encuentro en este contexto, por ejemplo, cierta DMU aumenta sus outputs un 5%? ¿Varía el resultado? Si es que no, continuamos con otro escenario hasta que detectamos el mínimo incremento necesario en los recursos para que el modelo clasifique a la DMU con la etiqueta contraria. De esta forma, obtenemos múltiples posibles escenarios y sabemos cómo responderá el modelo. De todos los futuros posibles, nos interesa aquel que, con el mínimo incremento, o reducción, de los recursos, el modelo cambia de decisión.

Ese incremento la definimos como ineficiencia técnica.

El score de eficiencia que proporciona nuestro método mediante aprendizaje automático buscamos que sea más robusto y obtener un análisis de la importancia de las variables más acertada.

Diap 7

Para explicar nuestro método, voy a presentar un ejemplo. Mediante un proceso generador de datos, generamos 30 unidades de toma de decisiones con un input que produce un output.

Diap 8

Para aplicar una técnica de clasificación hace falta tener datos clasificados. En contextos como la medicina, existen unos parientes, se registran sus datos, y al paso del tiempo, un médico valora y determina si los pacientes han desarrollado la enfermedad. Los datos ya estarían clasificados y se puede entrenar el modelo y comparar la predicción con la verdadera etiqueta que ocurrió.

En nuestro contexto, el experto es el DEA. El objetivo del DEA es determinar una tecnología basada en los datos y en ciertos axiomas para estimar una función de producción, no decreciente, convexa y determinista, o frontera de mejores prácticas. Si una DMU está en la frontera de mejores prácticas, la etiquetamos como eficiente, si no como ineficiente.

Diap 9 Step 1

En concreto, para el proceso de etiquetado, empleamos el modelo aditivo, con el objetivo de no asignar como eficientes aquellas observaciones que son débilmente eficientes. Solo las que son Pareto-eficientes.

De las 30 observaciones, 4 son Pareto-eficientes y las clasificamos como tal.

Podríamos entrenar el modelo directamente, pero existe un problema. Los datos están desbalanceados. Esto puede hacer que las métricas de rendimiento del modelo indiquen una precisión y un porcentaje de acierto de casi el 100%, cuando en realidad clasifica cualquier observación como ineficiente. Es un error asumible.

Para solucionar esto, existen diferentes métodos en la literatura. Nosotros proponemos la generación sintética de datos como método de sobrerrepresentar a la clase minoritaria, que consiste en crear una nueva observación que sea una combinación convexa de dos observaciones minoritarias.

Diap 10

En vez de generarla aleatoriamente, preguntamos a nuestro experto. Las proyecciones, es decir, las unidades que observamos en la frontera de mejores prácticas, son una combinación convexa, así que usamos esas proyecciones, basadas en datos observados, para aumentar la proporción de la clase eficiente.

Si hay 4 eficientes y 26 ineficientes, tras el balanceo, pasamos a 26 eficientes y 26 ineficientes.

Sin embargo, puede pasar que con ciertas kernels en SVM u otros métodos de clasificación, consideren que todo lo que no esté contenido en las etiquetas ineficientes, sea eficiente y el modelo prediga como eficientes valores extremos, que sabemos que son muy ineficientes.

Por esta razón, le introducimos más información al modelo.

Diap 11

Empeoramos aleatoriamente todas las observaciones entre en valor observado y el mínimo en cada una de las dimensiones, buscando que, durante el entrenamiento, por no fallar, elija los hiperparámentros que tienen más sentido económicamente.

Diap 12

Tras este tratamiento, pasamos de un dataset de 30 DMus a 82 dmus, de las cuales 26 son eficientes y 56 son infiencientes.

Entrenando un modelo SVM con una kernel polinómica y varios hiperparámetros, tras emplear validación cruzada,

Diap 13

Los hiperparametros elegidos es que el grado del kernel polinomico es 3, y el coste es 1.

Al realizar en entrenamiento, calculamos el número de errores de predicción respecto a lo etiquetado por el DEA en 100 escenarios, asignando en cada uno un umbral distinto para considerar a una observación como eficiente. Aquel escenario con menos diferencias respecto al experto, que falla menos, es el umbral que indicamos.

Para este caso con pocos datos es muy elevado. Solo las observaciones que tienen una probabilidad superior al 82% de ser efiecientes son considerados como como tales.

Para cada DMU, realizamos el análisis contrafactual buscando el cambio en la etiqueta, y cuando la probabilidad supera el umbral, determinamos el score de eficiencia como la media entre los dos últimos incrementos.

Realizamos a continuación un análisis de sensibilidad para conocer la importancia de las variables en el modelo, indicando una importancia de un 52,1% para el output y un 47,9% para el input.

Diap 14

Vamos a mostrar nuestro método de clasificación con datos reales. Para ello, utilizamos la base de datos PISA 2018 para 999 colegios españoles, que ofrece la OCDE, que evalúa las competencias de los estudiantes y de los centros.

Utilizaremos como variables EDUQUAL que indica la calidad de los recursos físicos educativos, ESCS es un índice de estado socioeconómico de las estudiantes, TSRATIO es el ratio profesor-estudiante. OUTPUT. Además, tenemos dos variables de contexto. Region, que es la comunidad autónoma a la que pertenece el centro y el tipo de colegio, publico, concertado o privado.

Diap 15

Diap 16

A este conjunto de datos de 999 observaciones, el modelo aditivo solo identidfica como eficietnes a 38 colegios, un 3,8% del total. Tras nuestro tratamiento para mejorar las proporciones, obtnenemos 961 colegios eficientes y 1960 colegios ineficientes, 33 y 67 respectivamente.

Aplicamos tanto SVM con una kernel polinómica con un conjunto de hiperparametros a probar como una red neuronal con una capa oculta.

Para observar los resultados, hemos representado las funciones de densidad de los scores y observamos que las funciones resultantes de aplicar las técnicas de aprendizaje automático son muy similares y que se diferencian del enfoque tradicional de DEA. Las técnicas de aprendizaje automático obtienen una media inferior al método estándar y detectan más observaciones como eficientes.

Además, el clasificador llega a clasificar con la etiqueta contraria a ciertas observaciones, especialmente SVM.

La red neuronal proporciona un score y SVM hay 9 observaciones infactibles, mismo número que detecta el DEA.