

CÁTEDRA: Aprendizaje Automático I

TECNICATURA SUPERIOR EN CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL

ALUMNA:

Cader Fernanda

DOCENTE: María Elena BAIN

Fecha de entrega :21 de Junio de 2023

ANA08: Caracterización de los algoritmos

Para resolver el problema de predicción de la incidencia de enfermedades cardíacas con el conjunto de datos de la Clínica Cleveland, se podría considerar una serie de algoritmos, entre ellos:

1. **Regresión logística:** Este es un buen punto de partida para problemas de clasificación binaria como este. La regresión logística es un método estadístico que busca modelar la probabilidad de un evento ocurriendo (en este caso, la presencia de enfermedades cardíacas) en función de otros factores.

2. **Árboles de decisión y Bosques Aleatorios:** Los árboles de decisión son intuitivos y fáciles de interpretar, lo que los hace útiles para identificar las características más importantes. Los bosques aleatorios, que combinan múltiples árboles de decisión, tienden a tener un rendimiento superior y pueden evitar el sobreajuste, un problema común con los árboles de decisión.

3. **Máquinas de vectores de soporte (SVM):** Los SVM pueden ser efectivos en espacios de alta dimensión y son adecuados para problemas de clasificación binaria. Sin embargo, pueden ser más difíciles de interpretar que los modelos de regresión logística o los árboles de decisión.

4. **Redes Neuronales:** Este tipo de algoritmo podría ser una buena opción si los otros modelos no son suficientes, ya que pueden capturar interacciones complejas entre las variables. Sin embargo, requieren más datos y más tiempo para entrenarse, y sus predicciones no son tan fácilmente interpretables.

5. **K-Nearest Neighbors (KNN):** Este es un algoritmo simple y fácil de entender. Funciona al asignar una nueva instancia a la clase que es más común entre sus K vecinos más cercanos. Sin embargo, KNN puede ser computacionalmente intensivo y no funciona bien con características irrelevantes.

6. **Gradient Boosting:** Este algoritmo crea un modelo predictivo en forma de un conjunto de modelos de predicción débiles, generalmente árboles de decisión. Funciona optimizando iterativamente cualquier función objetivo diferenciable arbitraria. Gradient Boosting es uno de los algoritmos más poderosos y efectivos para problemas de clasificación y regresión estructurados.

7. **Redes neuronales convolucionales (CNN) y Redes Neuronales Recurrentes (RNN):** Aunque son más comunes en el análisis de imágenes (CNN) y el procesamiento del lenguaje natural (RNN), estas redes neuronales también pueden ser útiles para tareas de clasificación.

Cabe destacar que el sobreajuste (overfitting) y el subajuste (underfitting) son problemas comunes en el aprendizaje automático que se deben vigilar. El sobreajuste ocurre cuando un modelo se ajusta demasiado bien a los datos de entrenamiento y no generaliza bien a nuevos datos. Por otro lado, el subajuste ocurre cuando un modelo no puede capturar la estructura subyacente de los datos.

En cuanto a las métricas, la curva ROC (Característica Operativa del Receptor) y el AUC (Área Bajo la Curva) son comúnmente utilizados para evaluar la eficacia de los modelos de clasificación. La curva ROC es un gráfico de la tasa de verdaderos positivos contra la tasa de falsos positivos para diferentes umbrales de decisión, y el AUC mide el área total bajo la curva ROC.

Finalmente, es esencial recordar que ninguna cantidad de algoritmos sofisticados puede compensar los datos de mala calidad. La ingeniería de características y la limpieza de datos deben ser las primeras etapas de cualquier proyecto de aprendizaje automático. Debemos asegurarnos de tener un buen entendimiento de tus datos y de cómo los algoritmos que estás utilizando funcionan para obtener los mejores resultados posibles.

Bibliografía:

Gèron, A (2017). Hands-On Machine Learning with Scikit-Learn and Tensor Flow. O'Reilly Media.

Grus, J (2015). Data Science from Scratch. O'Reilly Media.

Instituto de Tecnología de Santa Cruz. (2023). Unidad 3: Inteligencia Artificial. Recuperado el 13 de junio de 2023, de <https://inset.tecnicasantacruz.edu.ar/pluginfile.php/128963/mod_resource/content/5/AA1_Unidad3.pdf>

Norman, Alan T. (2017). Aprendizaje automático en acción.