

Proyecto de Programación III.

Miguel Ángel Jiménez Avendaño, Clara Isabel Botero Pérez.

23 de noviembre del 2024.

Ingeniería Biomédica.

Universidad EIA.

Modelo 1.

En el modelo 1 se utilizó regresión robusta, que es una técnica utilizada para estimar los parámetros de un modelo de regresión en presencia de valores atípicos. Entrenamiento, error variado. En el entrenamiento, el error (curva azul) es reducido y alcanza estabilidad de forma veloz al incrementar el tamaño de los datos de entrenamiento. Esto señala que el modelo captura de manera eficiente la conexión entre las características y el objetivo en los datos que analiza. La curva se mantiene estable, lo que indica que el modelo no está experimentando fluctuaciones significativas y no está siendo severamente afectado por el sobreajuste.

En la curva de validación (verde), el error es mayor inicialmente que el de entrenamiento, pero la diferencia es razonable; el modelo tiene un buen balance entre sesgo y varianza, y no está sobreajustado a los datos de entrenamiento.

La curva de validación también se estabiliza, lo que indica que agregar más datos probablemente no reduciría significativamente el error.

Modelo 2.

El error de entrenamiento (azul) muestra que a medida que aumenta el tamaño del conjunto de entrenamiento, el modelo se ajusta cada vez mejor a esos datos, esto debido a que el modelo tiene más información para aprender.

Por otro lado, el error de validación (amarillo), al principio tiende a disminuir, lo que indica que el modelo está aprendiendo patrones generales que le permiten hacer buenas predicciones en nuevos datos.

La distinción entre las dos líneas simboliza el sobreajuste. Si la línea azul se encuentra considerablemente más próxima al eje x que la amarilla, indica que el modelo se ha ajustado demasiado a los datos de entrenamiento y no logra generalizar adecuadamente con nuevos datos.

Modelo 3.

Este modelo utiliza un SVR (regresión de vectores de soporte), que es un algoritmo de aprendizaje automático que se utiliza para problemas de regresión y clasificación; encuentra una función que se aproxime lo mejor posible a un conjunto de datos continuos.

El error de entrenamiento (rojo) disminuye a medida que aumenta el tamaño del conjunto de entrenamiento. Esto indica que el modelo está aprendiendo los patrones en los datos de entrenamiento.

Mientras tanto, el error de validación (verde) disminuye igualmente a medida que aumenta el tamaño del conjunto de entrenamiento, pero a un ritmo más lento que el error de entrenamiento. Esto da a entender que el modelo se está generalizando correctamente a datos no vistos.

La diferencia entre los errores de entrenamiento y validación sugiere cierto nivel de sobreajuste. Esto indica que el modelo está memorizando parte del ruido presente en los datos de entrenamiento, lo cual puede resultar en un rendimiento deficiente en nuevos datos. En general, la trama indica que el modelo SVR es capaz de aprender de los datos y generalizar de manera efectiva a datos no vistos.

Comparación de métricas entre modelos.

En la gráfica de comparación de métricas entre el modelo 1 y 2 se analiza la precisión de un modelo al realizar pronósticos. En el eje horizontal se encuentran tres métricas habituales en la evaluación de modelos de regresión: el error cuadrático medio (RMSE), que calcula la media de las discrepancias entre las predicciones del modelo y los valores reales; cuantos valores más bajos se obtengan, mejor será la adaptación del modelo. El MAE (error absoluto medio) se basa en calcular el promedio de los valores absolutos de las diferencias. No se ve afectado fácilmente por valores poco comunes. R^2 , el coeficiente de determinación, refleja la parte de la variabilidad de los datos que queda explicada por el modelo. Cuando los valores se acercan a 1, esto indica que el ajuste es óptimo.

En el eje vertical se representan los valores de cada métrica correspondientes a cada modelo. El Modelo 1 presenta un RMSE considerablemente superior al del Modelo 2, lo que sugiere que, en términos generales, las predicciones de dicho modelo se encuentran más distantes de los valores reales. No obstante, registra un valor de MAE ligeramente inferior al del Modelo 2, indicando que, a pesar de presentar errores más significativos en ciertos casos, en general sus errores son más reducidos. El valor de R^2 es negativo, lo cual indica que el modelo no logra explicar la variabilidad de los datos y, en realidad, su rendimiento es inferior al de simplemente predecir el valor promedio.

Por otro lado, el Modelo 2 destaca por mostrar un RMSE significativamente inferior al del Modelo 1, lo cual sugiere que sus predicciones se acercan más a la realidad. El valor de MAE es incluso más bajo que el del Modelo 1, lo que confirma su mayor precisión en general. El coeficiente R^2 es ligeramente positivo y se acerca a cero, lo que indica que el modelo explica una cantidad mínima de la variabilidad de los datos.

En resumen, el Modelo 2 muestra un rendimiento superior al del Modelo 1. Presenta un error cuadrático medio y un error absoluto medio reducidos, lo cual sugiere que sus estimaciones son más acertadas.

Conclusión.

Se utilizaron estos 3 modelos, debido a que la base de datos tenía ciertas condiciones que no hacían viable usar los otros métodos.

El mejor modelo es el 3, el de SVR, ya que tiene una tendencia más estable entre los errores de entrenamiento y validación, indicando buen equilibrio entre capacidad de generalización y ajuste; presenta una curva consistente, con menor diferencia entre errores.