

Materia: Aprendizaje Automatico

Actividad: clase 4

Ejercicio 1 -

Regresión múltiple: Presión arterial

En este primer ejercicio trabajamos con un dataset que incluía las variables edad, horas de ejercicio, peso, nivel de estrés, ingresos y horas frente a la televisión. El objetivo era predecir la presión arterial.

Fui probando distintas combinaciones de variables. Cuando use todas las variables juntas, el modelo funcionó de manera aceptable, con un R^2 cercano a 0.71. Sin embargo, al enfocarme solo en las variables relacionadas a lo físico (edad, peso, estrés y horas de ejercicio) el modelo incluso mejoró, llegando a un R^2 de 0.74 y un RMSE más bajo.

En cambio, cuando intente predecir la presión arterial solo con los ingresos y las horas frente a la TV, el resultado no fue representativo ya que se obtuvo un R^2 negativo que mostraba que esas variables no tienen relación significativa con la variable objetivo.

También probe combinaciones más pequeñas, como edad y peso, y si bien aportan algo de explicación ($R^2 \approx 0.52$), no alcanzan para tener un buen modelo.

En conclusión, las variables que realmente influyen en la presión arterial son las fisiológicas: edad, peso, estrés y ejercicio. Las socioeconómicas no resultaron relevantes en este análisis.

Ejercicio 2 -

Regresión lineal: PIB mundial

En el segundo ejercicio usamos un dataset con el PIB de 51 países entre los años 2019 y 2023. A partir de esos datos construí la serie del PBI mundial y tratamos de explicarla en función de los PBI de distintos países.

Al usar todos los países a la vez, el modelo dio un ajuste perfecto ($R^2 = 1$), pero esto no es realmente significativo, porque son demasiadas variables para tan pocos datos. Lo interesante fue probar con algunos países específicos.

Materia: Aprendizaje Automatico

Actividad: clase 4

Cuando use solo Estados Unidos o solo China, el modelo explicó muy poco (R^2 entre 0.04 y 0.08). En cambio, cuando combine Estados Unidos, China e India, el poder explicativo subió bastante, alcanzando un R^2 de 0.76. Esto muestra que las grandes potencias mundiales sí permiten capturar buena parte de la variación del PBI mundial.

Por otro lado, cuando probe con Argentina y Brasil, el modelo apenas alcanzó un R^2 de 0.14, lo que refleja que, aunque aportan algo, su influencia en el total mundial es mucho menor.

Puedo concluir que los países que son potencias económicas globales tienen mucho más peso a la hora de explicar el PBI mundial que países de menor participación en la economía global.

Ejercicio 3 –

Clasificación: Usuarios por sistema operativo

En el último ejercicio trabajamos con un dataset de usuarios que incluía las variables duración, páginas, acciones y valor. La variable objetivo era el sistema operativo de cada usuario: Windows (0), Mac (1) o Linux (2).

Entrene un modelo de regresión logística multiclase y lo probe con un conjunto de test. El accuracy que obtuvimos fue cercano al 70%, lo que muestra un desempeño aceptable.

Al mirar la matriz de confusión note que el modelo funcionó muy bien con los usuarios de Linux, que siempre fueron clasificados correctamente. Con los usuarios de Windows también tuvo bastantes aciertos, aunque hubo algunos confundidos con Linux. En cambio, con los usuarios de Mac fue donde más problemas tuvo el modelo, ya que varios fueron mal clasificados, en su mayoría como Windows.

En conclusión, el modelo logra un rendimiento razonable, pero todavía queda espacio para mejorar. Una forma de hacerlo sería trabajar con un volumen mayor de datos y probar otros clasificadores que tal vez se adapten mejor a la distribución de estas clases.