Apuntes Modulo 5

Hector Manuel Gonzalez Villarreal

178679

Evaluación de modelo

Overfitting underfitting

Gridsearch

La evaluación de los módulos nos sirve como funcionaran en el mundo real

Partir el Dataset en Entrenamiento y pruebas

Al final usar todo el dataset

Train\_test\_split de sklearn,

Los parámetros son la data , el target, porcentaje de test, y la semilla

Las respuestas son arreglos xtraint, xtestm y ytrain y ytest

El tamaño de la muestra para entrenamiento y testing importa para el error de generalización

Para esto es bueno usar el corss\_val\_score

Donde el primer parámetro es el tipo de modelo, x, y y cv como particiones

Devuelve scores, y podemos usar np.mean() para saber cuanto es el score en promedio

Para predecir usarmos cross\_val\_predict, con los mismos parámetros

Pero esta vez entrega un arreglo de predicciones

Cuando el modelo es muy flexible y ajusta los resultados podemos decir que esta sobreentrenado

R cuadrada nos ayuda para ver cual es el grado del polinoimio correcto

Ridge regresión

Previene sobreentrenamiento

Intriduce un parámetro alfa para 0 a 1 partiendo de .001 y 0.01

Usamos ridge(alpfa=valor)

Hacemos un fit

Por cada valor y usamos el scores usamos es valor de alfa que maxifique la r2

Gridsearch

Busca paramtetros o hiperparametros para buscar las mejores

Para esto se hace un arreglo de lista de valores por cada llave

La salida es una matriz de scores con los diferentes parámetros

Lo mejos es el .best\_estimator

Y para ver todos los scores usamos cv-results\_