15-06-2021

Ensemble, tipos:

Secuencial: Votación.- utiliza la moda el hard voting. El soft voting trabaja con la media. Si es regresión nos quedamos con la media de los diferentes modelos de regresión

Distintos modelo,

Modo paralelo: Mismo modelo: bagging (distintas muestras con reemplazamiento) y pasting (distintas muestras sin reemplazamiento)

Bootstrap.- se busca que los estimadores sean lo más independientes posible

Bagging, cómo funciona

Boosting.- se ejecuta un algoritmo y con este entrenado vamos a entrenar el siguiente modelo y así sucesivamente

Adabost.- recoge el entrenamiento del anterior modelo

Grid search.-

Pipeline:

le damos los hiperparámetros que queremos para hacer el entrenamiento (quitar columnas, estandarizar, seleccionar las 20 columnas que mejor predicen, …)

Permite probar diferentes modelos con diferentes hiperparámetros, todos a la vez

cv.- cross validation

Simpleimputer.- asigna a los valores missing lo que le diga (una media, etc)

chi2 se le dice que se quede con las x columnas que mejor correlacionen. Nosotros le damos el x

njob=-1; que pruebe todos los núcleos del procesador

simpleImputer por defecto es la media

standarized