

Modelos ensemble

Los modelos ensemble son un enfoque en el aprendizaje automático que implica combinar las predicciones de varios modelos individuales para mejorar el rendimiento general del sistema. Estos modelos combinados (también conocidos como "ensembles") suelen superar a los modelos individuales al reducir la varianza, aumentar la estabilidad y mejorar la capacidad de generalización del modelo. Hay varios métodos de ensemble, pero los dos más comunes son el Bagging y el Boosting.

1. Bagging (Bootstrap Aggregating):

- En Bagging, se crean múltiples modelos (generalmente del mismo tipo) entrenándolos en conjuntos de datos ligeramente diferentes. Cada conjunto de datos se obtiene mediante un muestreo con reemplazo (bootstrap) del conjunto de entrenamiento original.
- El modelo final realiza una combinación de las predicciones de estos modelos base, a menudo tomando un promedio (en el caso de regresión) o realizando una votación (en el caso de clasificación).

Ejemplos de modelos ensemble basados en Bagging:

- **Random Forest** Un conjunto de árboles de decisión entrenados en conjuntos de datos bootstrap.
- Bagged Decision Trees: Conjunto de árboles de decisión individuales entrenados en conjuntos de datos bootstrap.

2. Boosting:

- A diferencia de Bagging, el Boosting se centra en mejorar la precisión del modelo, asignando más peso a las instancias mal clasificadas.
- Se construyen modelos secuenciales, y cada nuevo modelo se enfoca en corregir los errores cometidos por los modelos anteriores.
- Al final, se realiza una combinación ponderada de las predicciones de todos los modelos base para obtener la predicción final.

Ejemplos de modelos ensemble basados en Boosting:

- AdaBoost (Adaptive Boosting): Asigna pesos a las instancias mal clasificadas y ajusta modelos sucesivos para corregir esos errores.
- Gradient Boosting: Construye modelos secuenciales que minimizan el error residual del modelo anterior.
- **XGBoost** (eXtreme Gradient Boosting): Una implementación eficiente y escalable de Gradient Boosting.

Los modelos ensemble son populares porque pueden mejorar significativamente el rendimiento predictivo y generalmente son más robustos frente a sobreajuste en comparación con modelos individuales. Sin embargo, también pueden ser más computacionalmente intensivos debido a la necesidad de entrenar y mantener varios modelos.

Ensemble Bagging vs Boosting.ipynb