

Aprendizaje Automático

Departamento de Informática – UC3M

TUTORIAL 7 – Gradient Boosting Trees

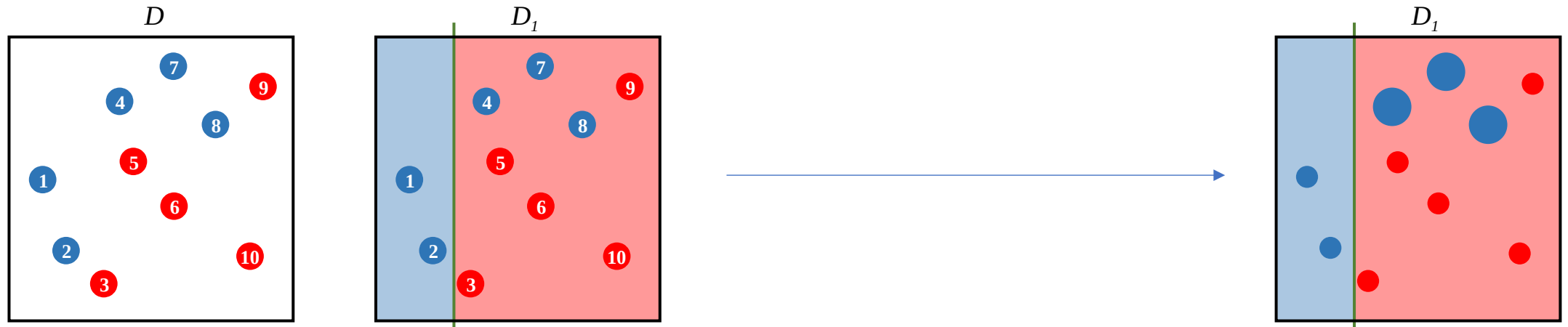
Tutorial 7

Recordando teoría. Boosting

Motivación: mejorar (*to boost*) métodos débiles (*weak learners*):

- Se usan modelos base que por sí solos consiguen un error no muy por encima del azar o *weak learners* (ej: árboles poco profundos)
- Se van añadiendo modelos base al *ensemble* secuencialmente, de tal manera que el siguiente modelo “corrija” los errores del modelo anterior

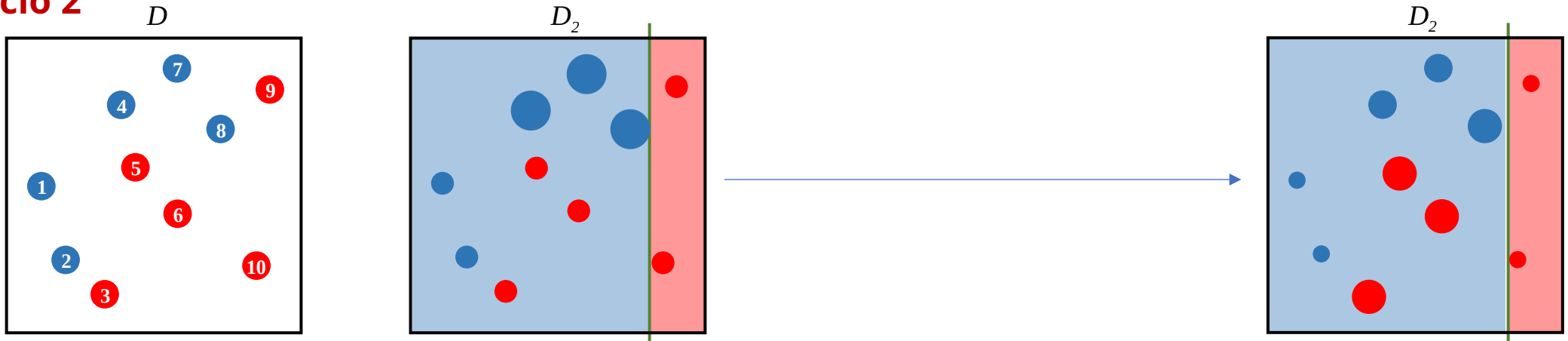
Ciclo 1



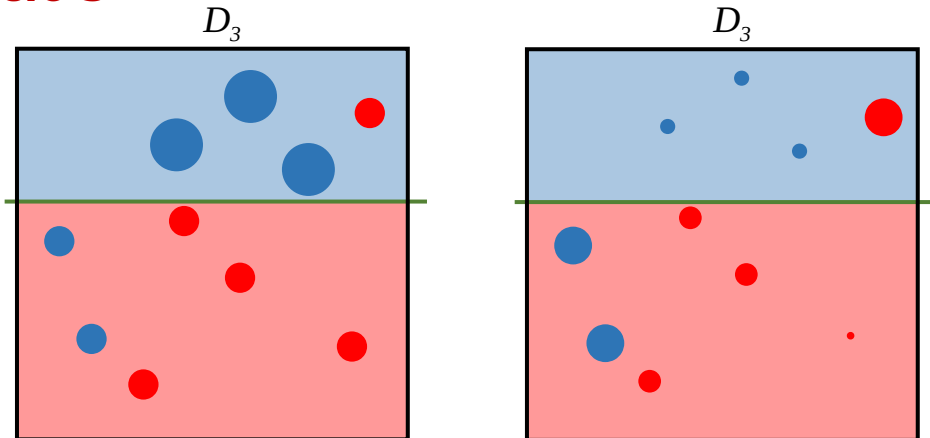
Tutorial 7

Recordando teoría. Boosting

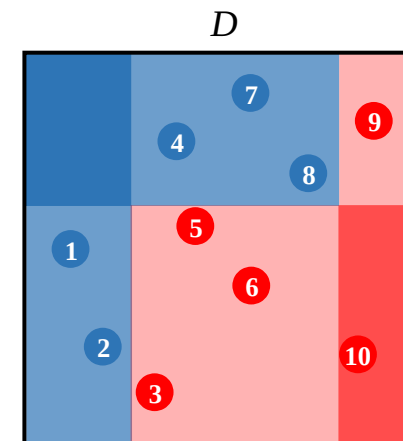
Ciclo 2



Ciclo 3



Resultado



Tutorial 7

Recordando teoría. Gradient Boosting

- Típicamente usa árboles como modelos base (*Gradient Boosted Tree*, GBT)
- Es habitual usar como función de pérdida
 - Para regresión: MSE
 - Para clasificación: $\log(LOSS)$
- Sigue el esquema *Boosting*
 - En cada iteración, ajusta modelos a nuevos datos, cuyas salidas son los errores del modelo anterior y la salida real

Tutorial 7

Recordando teoría. Gradient Boosting

• Hiperparámetros

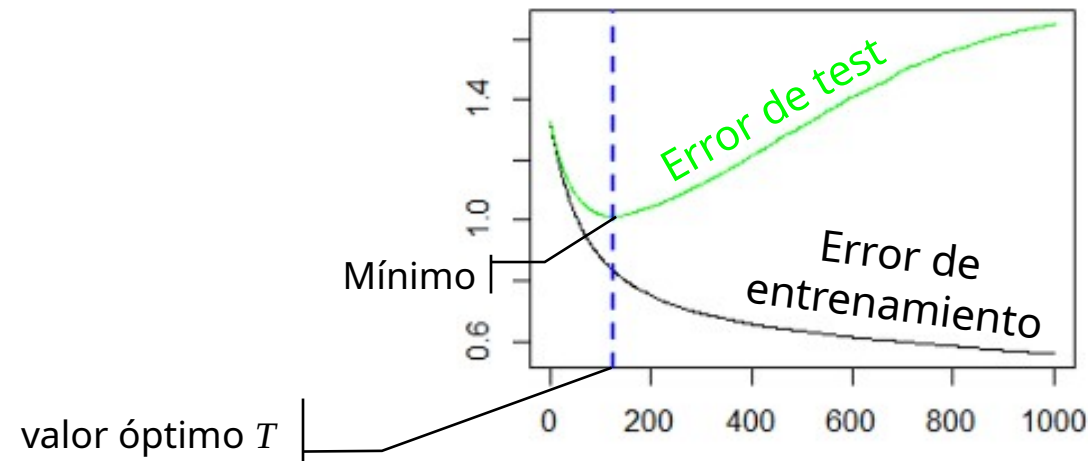
- Tasa de aprendizaje (*shrinkage*), ($0 < \nu < 1$)

$$f_T(\mathbf{x}) = f_{T-1}(\mathbf{x}) + \nu \alpha_{T-1} h_{T-1}(\mathbf{x})$$

- Hace que el nuevo modelo tenga menos peso y por tanto sobreajusta menos
- Si muy baja, convergencia lenta, coste computacional alto pero más robusto, por tanto mejor resultado, implica aumentar T

- Número de árboles del modelo, T

- Si convergencia lenta entonces hacen falta más árboles, pero muchos árboles (a diferencia de RF) producen sobreaprendizaje



Tutorial 7

Parada temprana. Gradient Boosting

- Con el número suficiente de weak learners, el modelo final tiende a ajustarse perfectamente a los datos de entrenamiento causando overfitting.
- Parada temprana (early stopping):
 - **validation_fraction**: proporción de datos separados del conjunto entrenamiento y empleados como conjunto de validación para determinar la parada temprana (early stopping).
 - **n_iter_no_change**: número de iteraciones consecutivas en las que no se debe superar el tol para que el algoritmo se detenga. Si su valor es None se desactiva la parada temprana.
 - **tol**: porcentaje mínimo de mejora entre dos iteraciones consecutivas por debajo del cual se considera que el modelo no ha mejorado.

Tutorial 7

Librerías

- Scikit-learn - HistGradientBoostingRegressor
- Otras librerías:
 - XGBoost (*Extreme Gradient Boosting*)
 - LightGBM

Tutorial 7

Random Forest vs Gradient Boosting

- RF tiene *out-of-bag error* y no necesita validación cruzada.
- RF tiene menos hiperparámetros
- Si tenemos muchos predictores irrelevante, RF puede tener menos prestaciones (selección de predictores).
- RF, cada modelo del ensemble es independiente y se puede paralelizar.
- Con una buena optimización de hiperparámetros GB suele tener mejores resultados.
- GB suele ser más rápido prediciendo y su tamaño suele ser menor.