Técnicas Machine Learning 1

Iván G. Torre

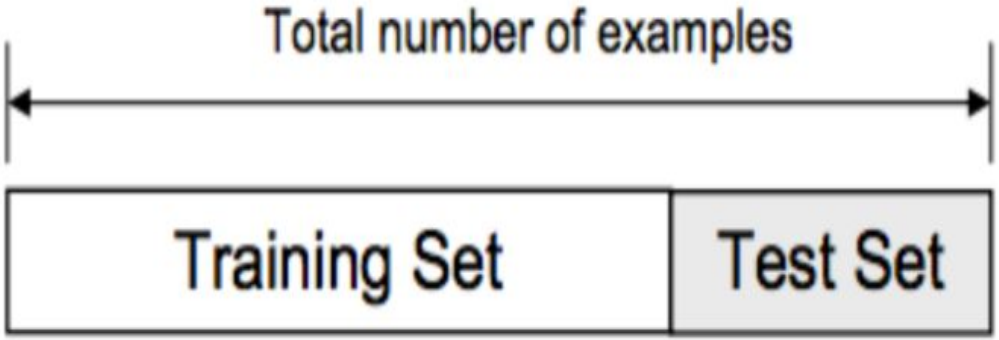
Iván G Torre: ivan.gonzalez.torre@upm.es

Jorge Calero: jorge.calero.sanz@alumnos.upm.es

José Olarrea: jose.olarrea@upm.es

División entre Train y Test en Clasificación

- Train: Model learn
- Test: Prediction

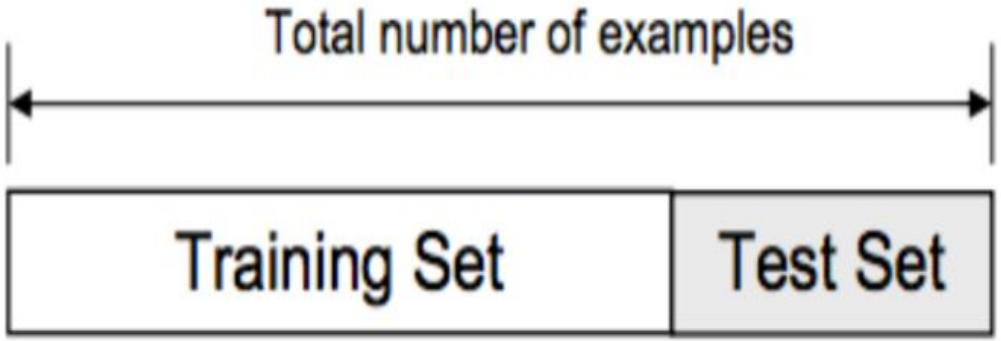


Train y Test: Problema

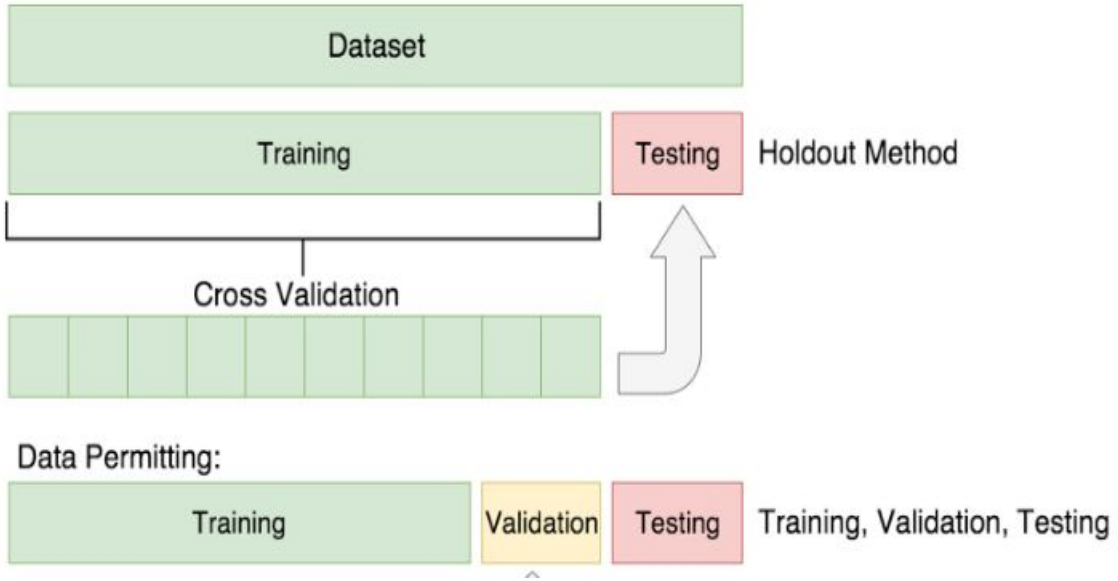
- Datos ordenados

Train y Test: Punto de división Train-Test

- Arbitrario
- Overfitting



Train y Test: Validación Cruzada

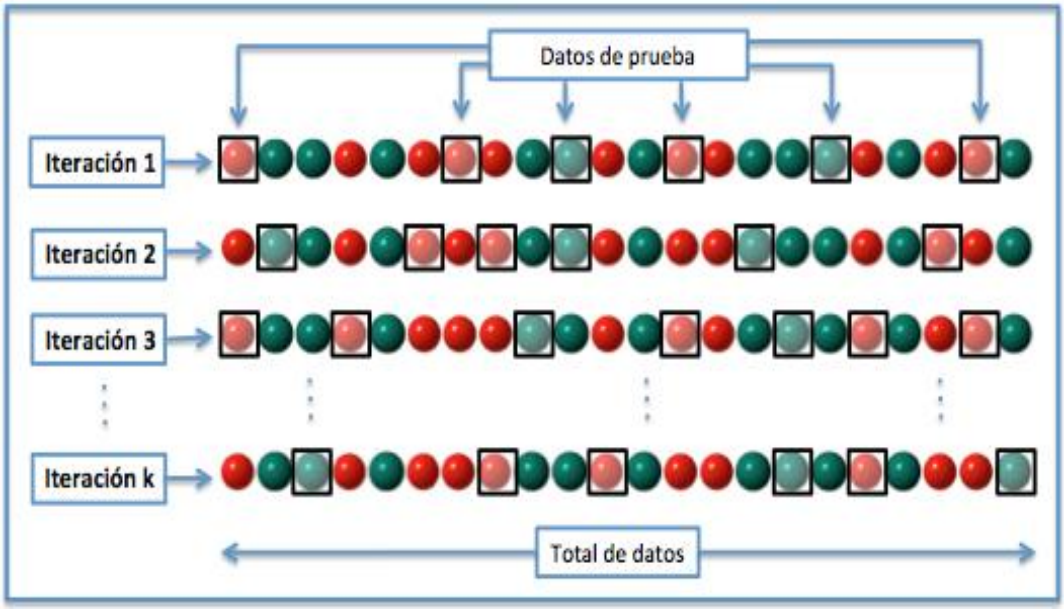


Train y Test: Validación Cruzada

- K-Fold:
  - Se divide en k subconjuntos
  - Uno es validation, (k-1) train

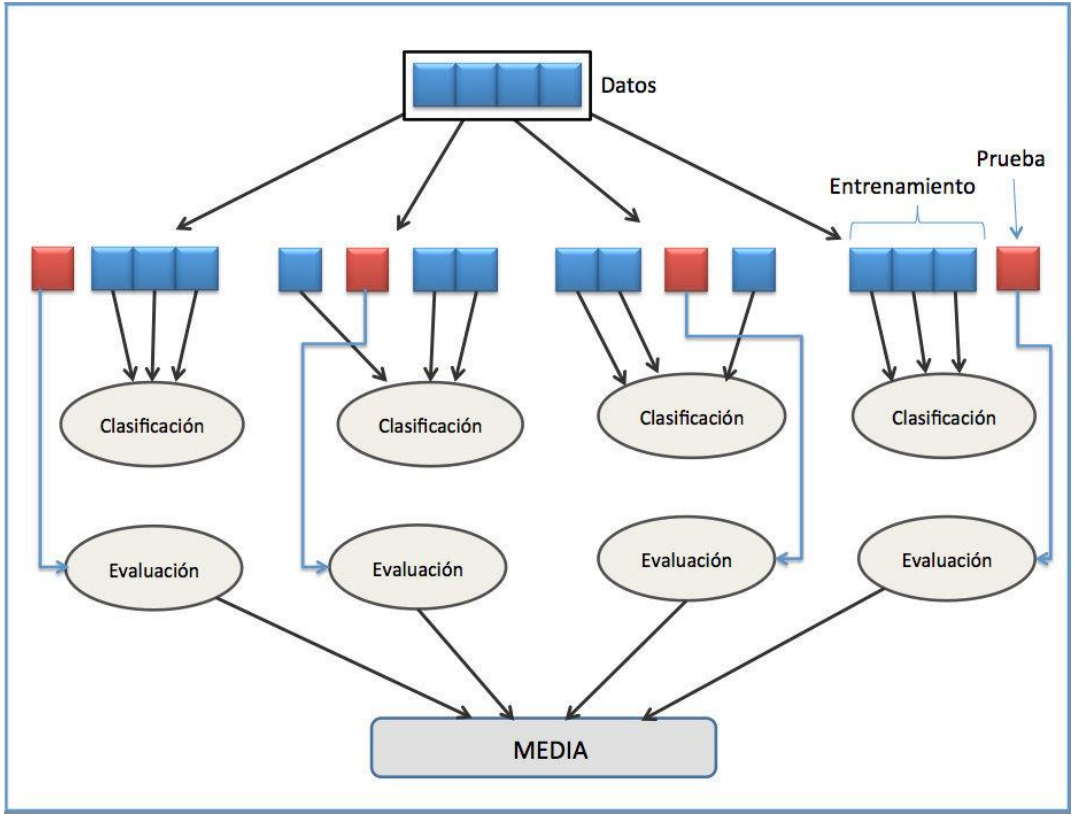
Train y Test: Validación Cruzada

- CV aleatoria:
  - Se elige aleatoriamente los datos de validación, el resto es train
- No depende del número de iteraciones
- Puede haber repetición



Train y Test.

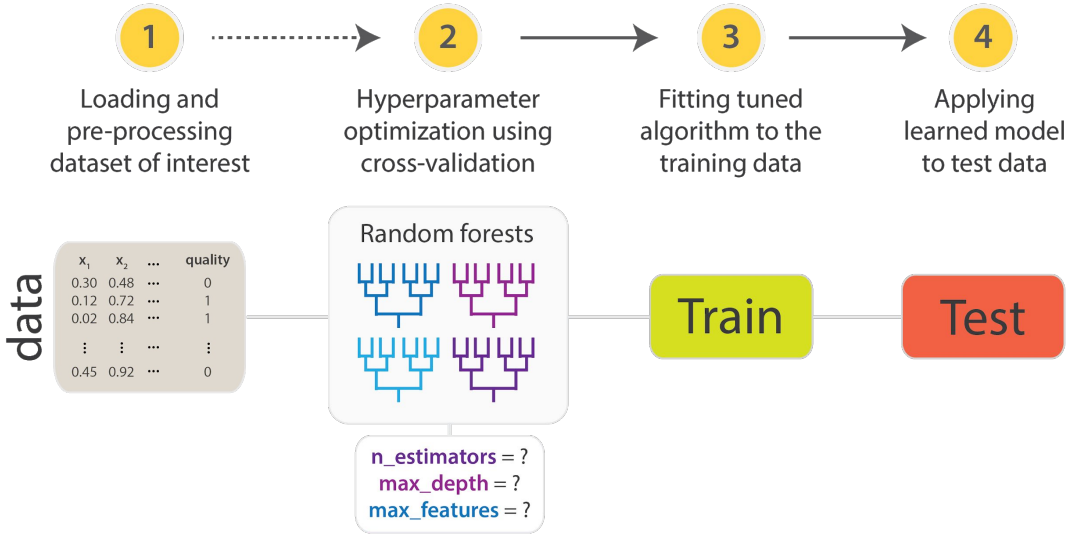
- CV uno fuera:
  - Muy costoso computacionalmente



Cross Validation: Ejercicio

Hyperparameter optimization

- Datos -> OK
- Elección Modelo -> OK
- Tuneado del modelo ->

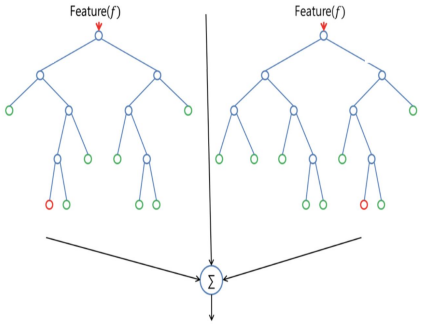


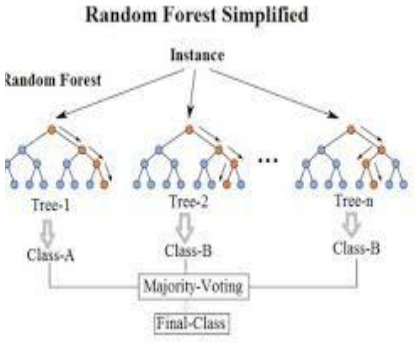
Hyperparameter optimization: Random Forest

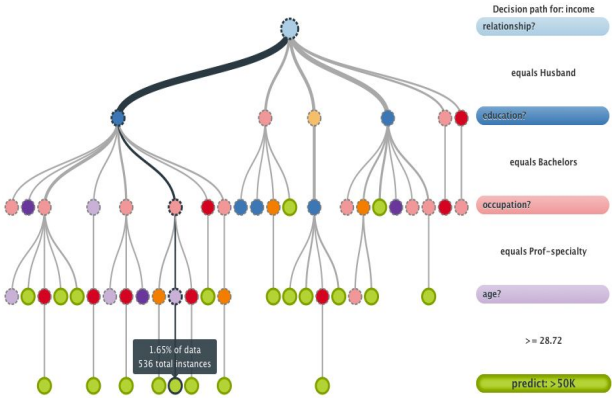
- **n_estimators : integer, optional (default=10)**
  - The number of trees in the forest.
- **max_depth : integer or None, optional (default=None)**

  The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.

- **n_jobs : int or None, optional (default=None)**
  - The number of jobs to run in parallel for both fit and predict. None means 1 unless in a joblib.parallel_backend context. -1 means using all processors. See Glossary for more details.
- **class_weight : dict, list of dicts, “balanced”, “balanced_subsample” or None, optional (default=None)**

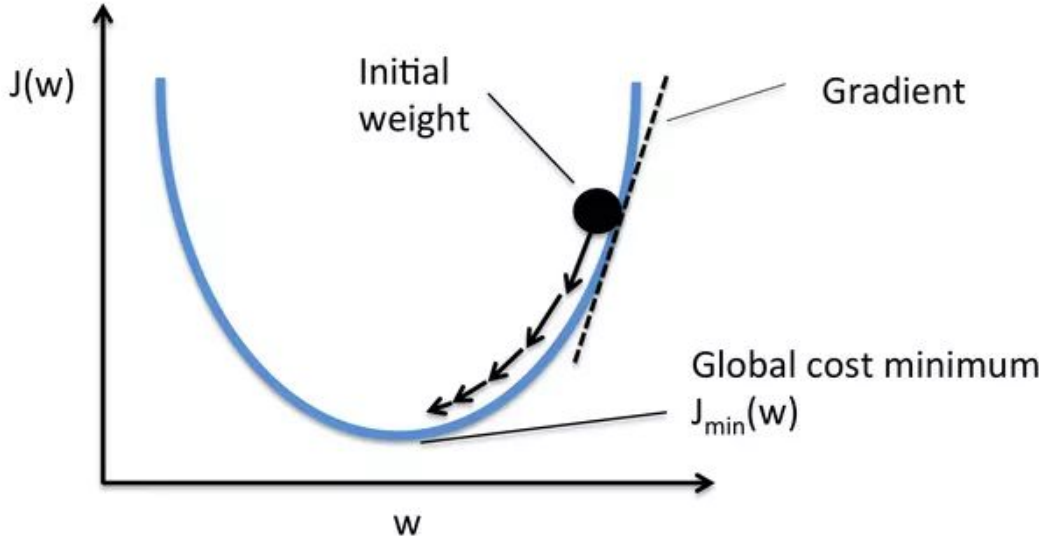






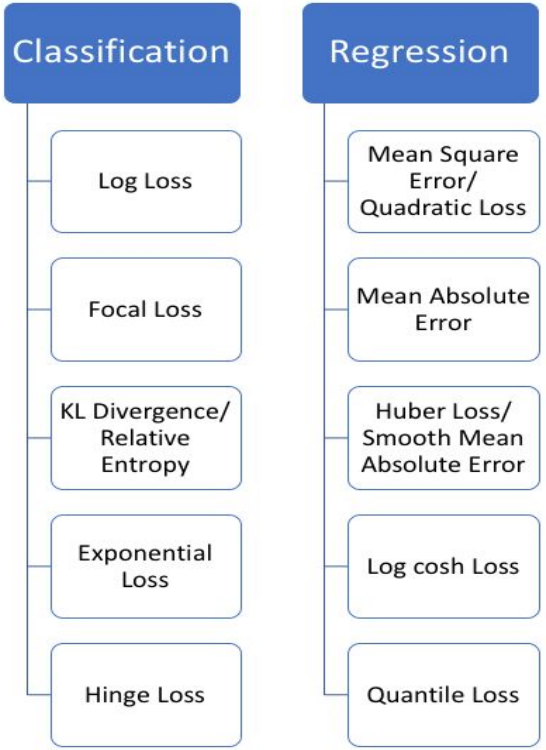
Hyperparameter optimization: loss function

- Necesitamos una métrica de error que minimizar



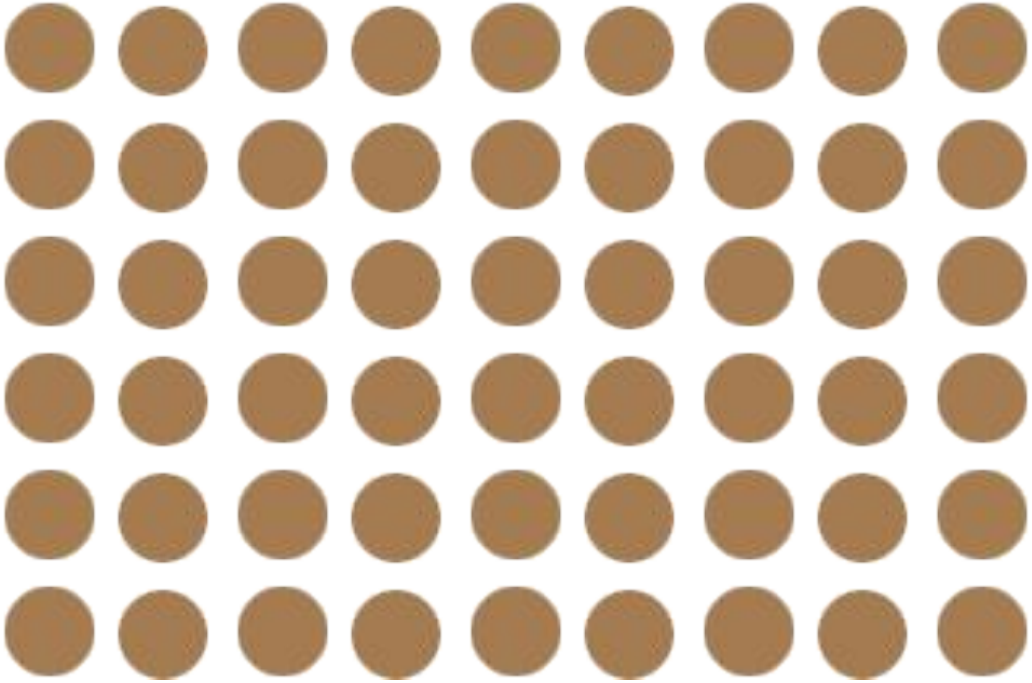
Hyperparameter optimization: loss function

- Dependiendo del modelo, existen muchas.
- Problema muy desarrollado para cálculo numérico entre otros



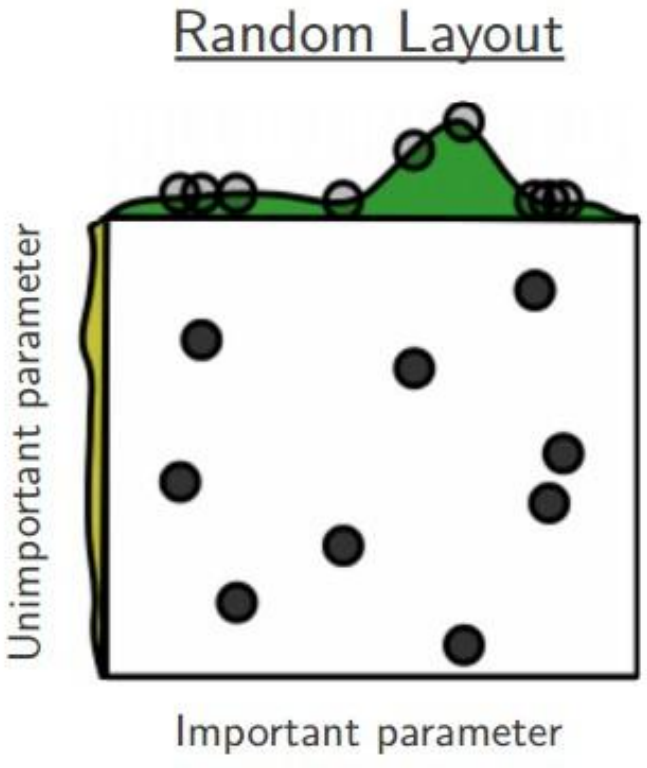
Hyperparameter optimization: Grid Search

- Búsqueda exhaustiva
- Paralelizable



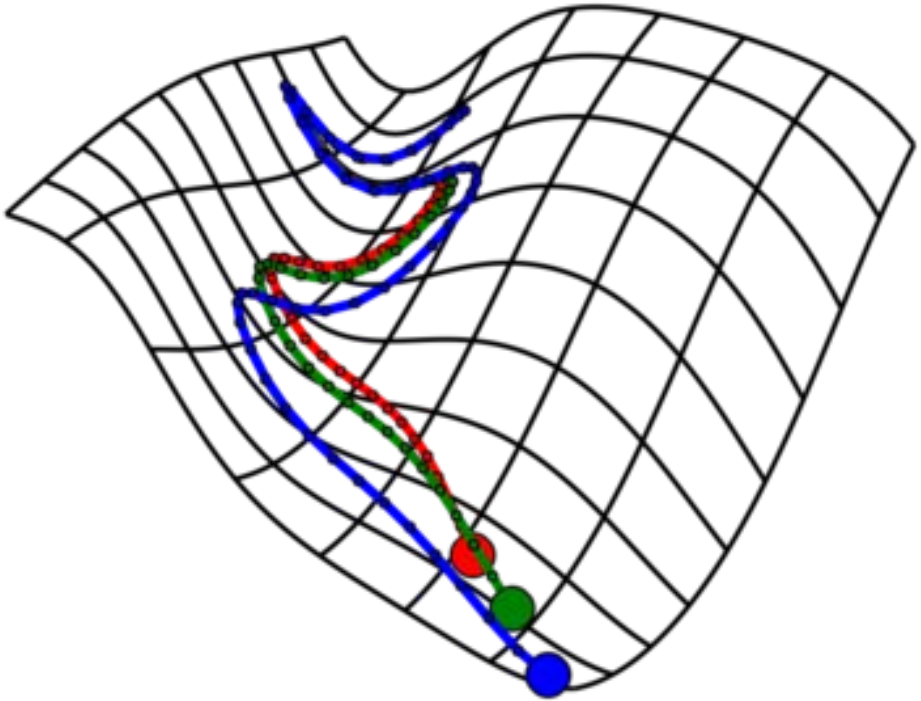
Hyperparameter optimization: Búsqueda aleatoria

- Búsqueda no exhaustiva
- Paralelizable



Hyperparameter optimization: Búsqueda gradiente

- Búsqueda no exhaustiva
- No Paralelizable
- Ampliamente estudiado



GridSearch en Clasificación: Ejercicio