



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Laboratorio de datos

Selección de modelos

Verano 2025

Manuela Cerdeiro - Paula Perez Bianchi - Viviana Cotik - Pablo Brusco (AA)

Evaluación de modelos \leftrightarrow **selección** de modelos

Necesitamos poder evaluar los modelos de una forma efectiva para:

- Comparar configuraciones de algoritmos
- Estimar la performance que tendrá el modelo “en la realidad”

Evaluar bien significa entender cómo será el uso, cuál es el objetivo del modelo, qué métrica refleja bien lo que queremos medir.

Selección de modelos

¿Por qué tendríamos distintos modelos para comparar?

- Distintos **atributos** (selección y transformación de atributos)
- Distintos **algoritmos (*modelos*)** (umbral, árboles, KNN, SVM, ...)
- Distintos **hiperparámetros** de cada algoritmo.

Ejemplo: hiperparámetros de los árboles de decisión

- Criterio de elección de atributos en cada nodo (Information Gain, Gini Gain...)
- Criterio de parada (ej: máxima profundidad)
- Estrategia de poda

Performance de un modelo - ¿dónde?



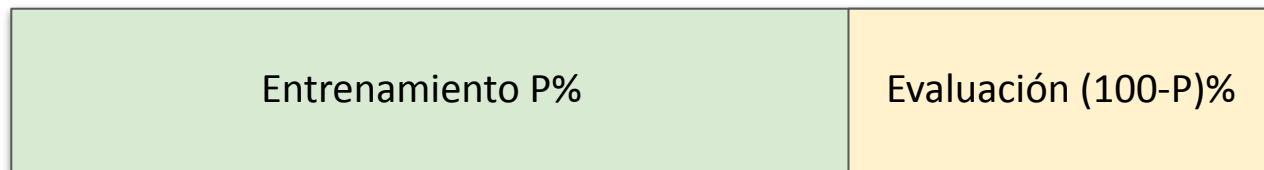
Medir la performance sobre datos de entrenamiento no es una buena idea. Surge la necesidad de separar un % de datos, para validar los modelos: datos de validación (o test).

Validación cruzada

Entrenamos nuestro modelo con **algunos** de nuestros datos, y vemos cómo funciona en los otros datos.

- al azar (elijo algunos, evalúo en otros)

¿Posibles problemas?

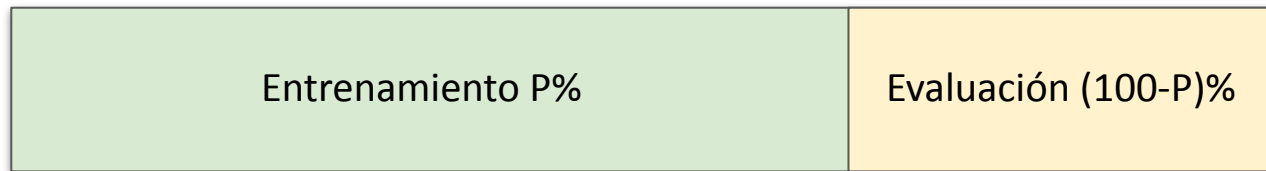


Validación cruzada

Entrenamos nuestro modelo con **algunos** de nuestros datos, y vemos cómo funciona en los otros datos.

- al azar (elijo algunos, evalúo en otros)
- al azar varias veces y promediar

¿Posibles problemas?



xN

Validación cruzada

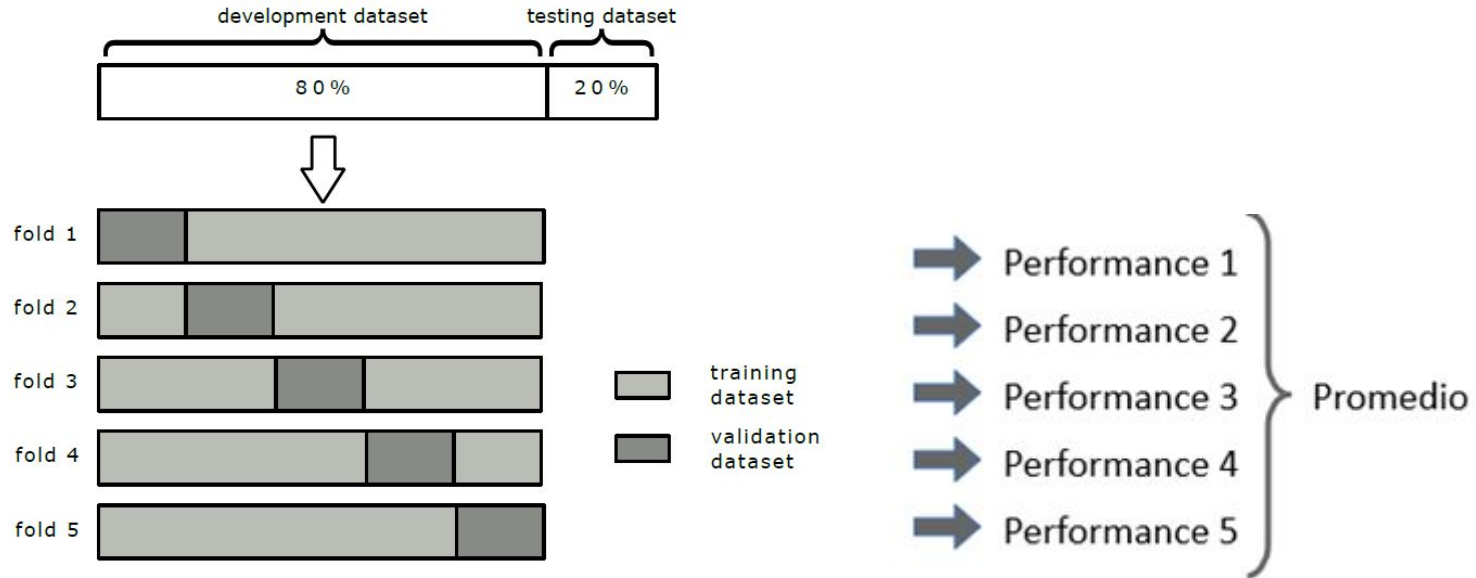
Entrenamos nuestro modelo con **algunos** de nuestros datos, y vemos cómo funciona en los otros datos.

- al azar (elijo algunos, evalúo en otros)
- al azar pero repitiendo varias veces (y promediando)
- leave one out
- k-folding

K-Fold cross-validation

1. Desordenar los datos (o no).
2. Realizar una partición del conjunto en k folds disjuntos del mismo tamaño.
3. Para $i = 1 \dots k$:
 - entrenar considerando todos los folds menos i .
 - predecir en el fold restante y medir la performance.
4. Promediar las métricas.

K-Fold cross-validation



Selección de modelos

Modelo: Cómo planeo modelar el proceso.

Por ejemplo, ¿árbol de decisión o SVM?

Hiperparámetros: Valores que se especifican previo a realizar el aprendizaje, es decir determinan el proceso de aprendizaje automático.

Por ejemplo, en árboles: el criterio de elección de atributos en cada nodo (Gini Gain, Information Gain), cantidad de hijos permitidos (árboles binarios vs. n-arios), criterio de parada (ej: max_depth), estrategia de poda, etc.

Parámetros: Valores internos del modelo resultante luego del aprendizaje, que se ajustan a partir de ejecutar el algoritmo sobre un dataset. Representan las reglas aprendidas.

Por ejemplo, en árboles: las preguntas a realizar en cada nodo del árbol.

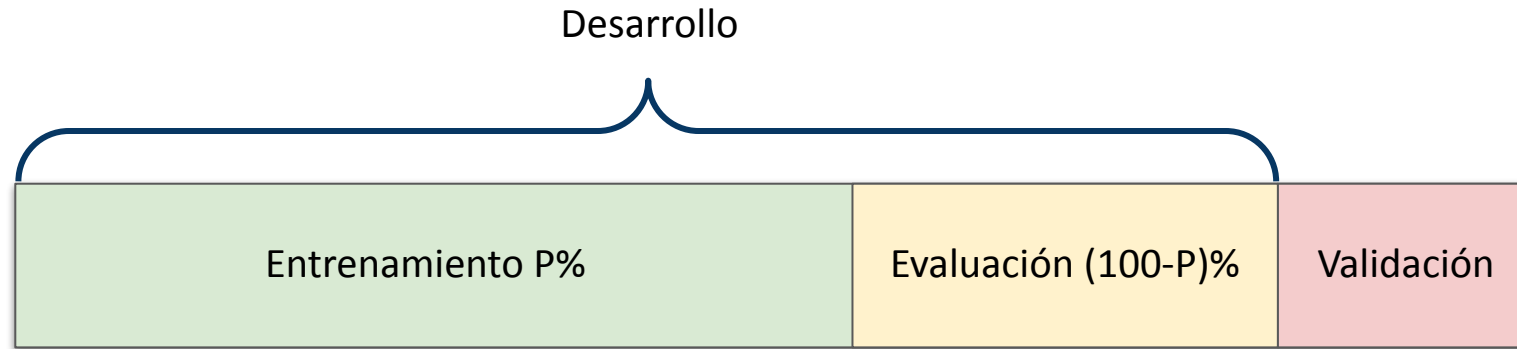
Selección de modelos con validación cruzada

Al realizar validación cruzada estamos entrenando k veces un **mismo modelo** con mismos hiperparámetros, ajustado en **distintos conjuntos**, lo que da lugar a distintos parámetros. Y luego promediamos la performance, para tener una idea de la **performance** de la elección modelo+hiperparámetros.

Cuando hablamos de selección de modelos, pensamos en **comparar** un mismo modelo pero configurado y ajustado con distintas configuraciones de hiperparámetros, o quizás (menos usual) distintos modelos. Si estamos utilizando cross-val, evaluamos todos los modelos en la misma partición.

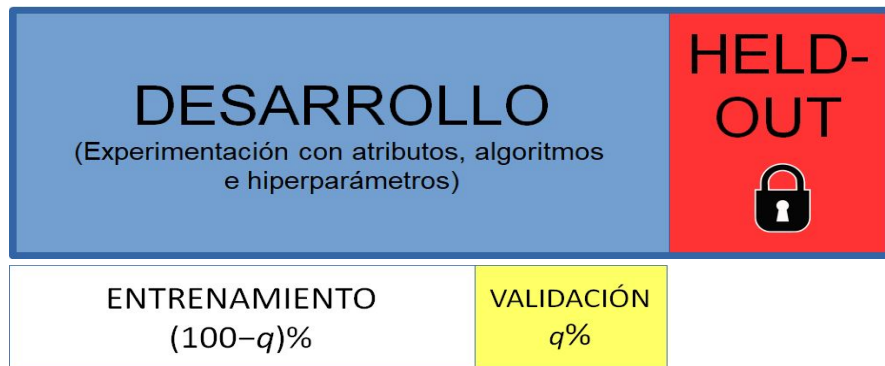
Validación cruzada

En realidad... si después queremos reportar la performance de nuestro mejor modelo.



Conjunto held-out (o control o test)

- Al comenzar hay que **separar un conjunto de datos** (Held-Out, Control o **Test**) y **NO TOCARLOS** hasta el final
- Todas las pruebas y ajustes se realizan sobre el conjunto de **Desarrollo**
- Al terminar todas las pruebas, se evalúa el modelo obtenido con el conjunto Held-out



Ejercicios: validación cruzada para clasificación

- Utilizar los datos 'seleccion_modelos.csv'. Se trata de un problema binario que queremos modelar utilizando un árbol de decisión.
- Separar un 10% de datos para el test final (held-out).
- Con los datos de desarrollo queremos utilizar k-fold cross-val para comparar distintas elecciones del hiperparámetro de altura, es decir 'max_depth'.
- Hacer k-folding con $k = 10$, probar con alturas 1, 2, 3, 5, 8, 13, 21. Determinar cuál fue la mejor opción. Anotar la altura seleccionada y el score obtenido.
- Utilizando la altura seleccionada, entrenar un árbol de decisión ajustando al conjunto entero de los datos, evaluarlo en el mismo conjunto total de datos y anotar este valor.
- Probar el modelo recién entrenado en el conjunto held-out y reportar su performance.

Ejercicios: validación cruzada para regresión

- Utilizar los datos de autos - mpg. Vamos a modelarlo con knn.
- Separar un 10% de datos para el test final (held-out). Los restantes serán los datos de desarrollo.
- Con los datos de desarrollo queremos utilizar k-fold cross-val para comparar distintas elecciones del hiperparámetro k (vecinos).
- Hacer k-folding con $k = 10$ (10 folds), probar valores de vecinos entre 1 y 50. Determinar cuál fue la mejor opción. Anotar el valor elegido y el score obtenido.
- Utilizando el valor elegido para el hiperparámetro, entrenar ajustando al conjunto entero de los datos, evaluarlo en el mismo conjunto total de datos y anotar este valor.
- Probar el modelo recién entrenado en el conjunto held-out y reportar su performance.

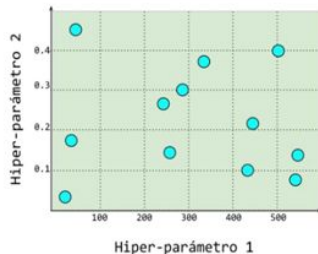
Selección de modelos

¿Cómo buscar la mejor combinación de **modelo + hiperparámetros**?

- Exploramos un espacio de búsqueda, **usando k-fold CV** para medir el desempeño de cada combinación.

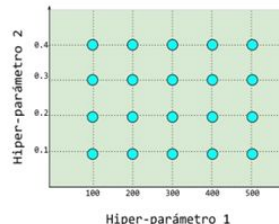
Random search (best guess, 1 factor at a time) Explorar opciones y combinaciones al azar

Ejemplo con 2
hiperparámetros



Grid search

Plantear opciones y explorar todas las combinaciones



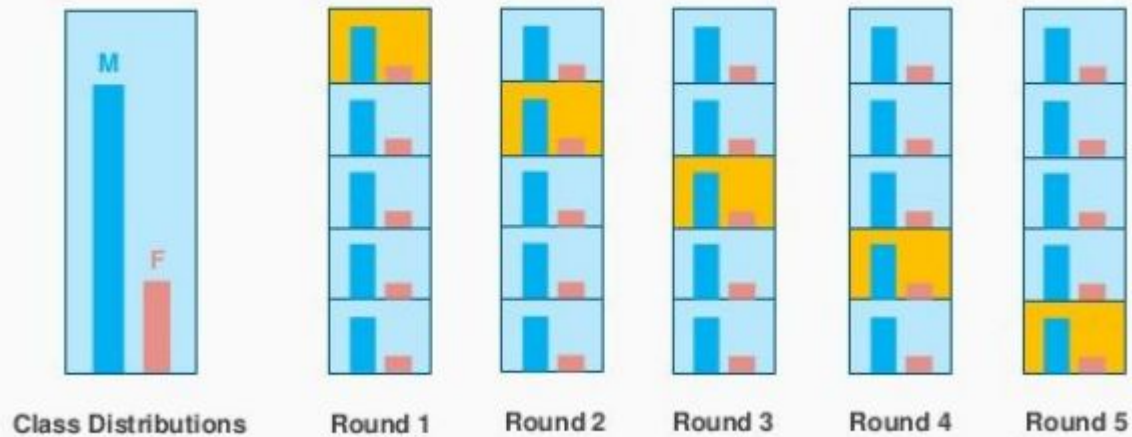
Al terminar, nos quedamos con la combinación con **mejor desempeño**,
y **entrenamos un único modelo usando todos los datos**

K-Fold cross-validation

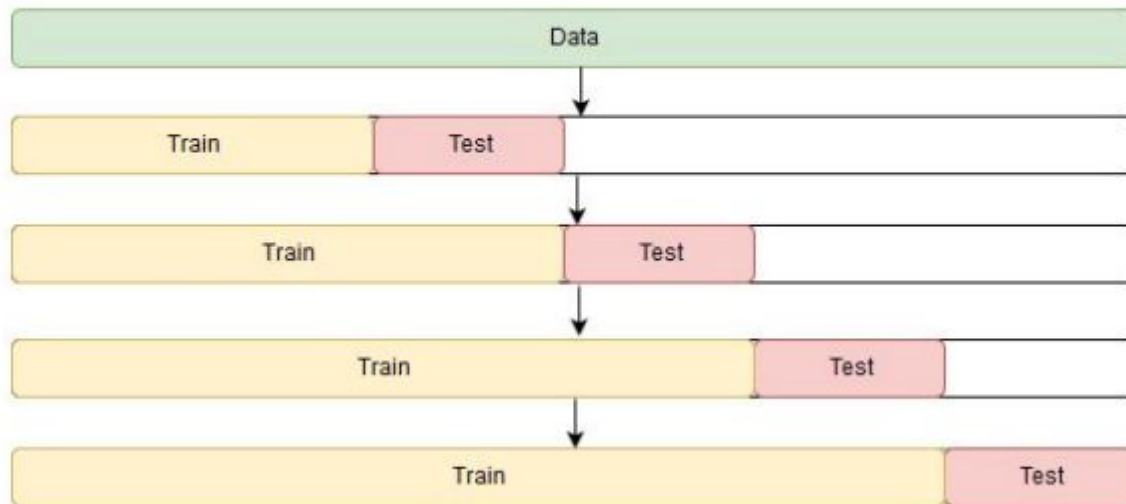
Algunas preguntas

- Es necesario mezclar al azar?
- Es bueno mezclar al azar?
- Cómo están ordenados los datos?
- Cómo están balanceadas las clases?

Stratified K-fold cross validation.



Temporal series K-fold cross validation.



Ejercicios

Para los experimentos realizados antes, probar las siguientes variantes y repetirlo.

- Probar realizar el experimento con otro valor de `N_folds` para el k-folding.
- Probar variando más de un hiperparámetro a la vez: por ejemplo en el árbol de decisión, `max_depth` y `criterion`.

Bibliografía

- Müller & Guido, "Introduction to Machine Learning with Python", O'Reilly, 2016.