

L U M B R E



División train/validation/test

Grado: 5

Materia: Fundamentos de Machine Learning · Machine Learning

03/10/2026

Antes de empezar

El error en los datos con que entrenaste no dice nada sobre el mañana. La división en tres conjuntos es el experimento controlado que permite imitar el futuro: aprender con uno, afinar con otro y juzgar con un tercero que el modelo jamás vio.

QUÉ APRENDERÁS

- Separar los datos en conjunto de entrenamiento, validación y prueba para medir de forma honesta la generalización y afinar el modelo sin contaminar la evaluación final.

División train/validation/test

No te autoexamines con las respuestas delante

Si mides tu modelo con los datos que usó para aprender, es como examinar a un alumno con las respuestas ya memorizadas: sacarás un sobresaliente falso. El conjunto de prueba virgen es lo único que imita el mundo real.

VIDEO

Datos de entrenamiento, validación y prueba

Vídeo en español sobre cómo crear los tres conjuntos de datos y qué objetivo tiene cada uno.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=vdYzm4xC7mc>

Por qué hace falta separar

El objetivo es la generalización: cómo se comportará con datos nunca vistos. Como no los tenemos, reservamos una porción real de los que sí tenemos y la tratamos como si fuera el futuro. Sin esa reserva, toda estimación del rendimiento sería optimista e inútil.

El conjunto de entrenamiento

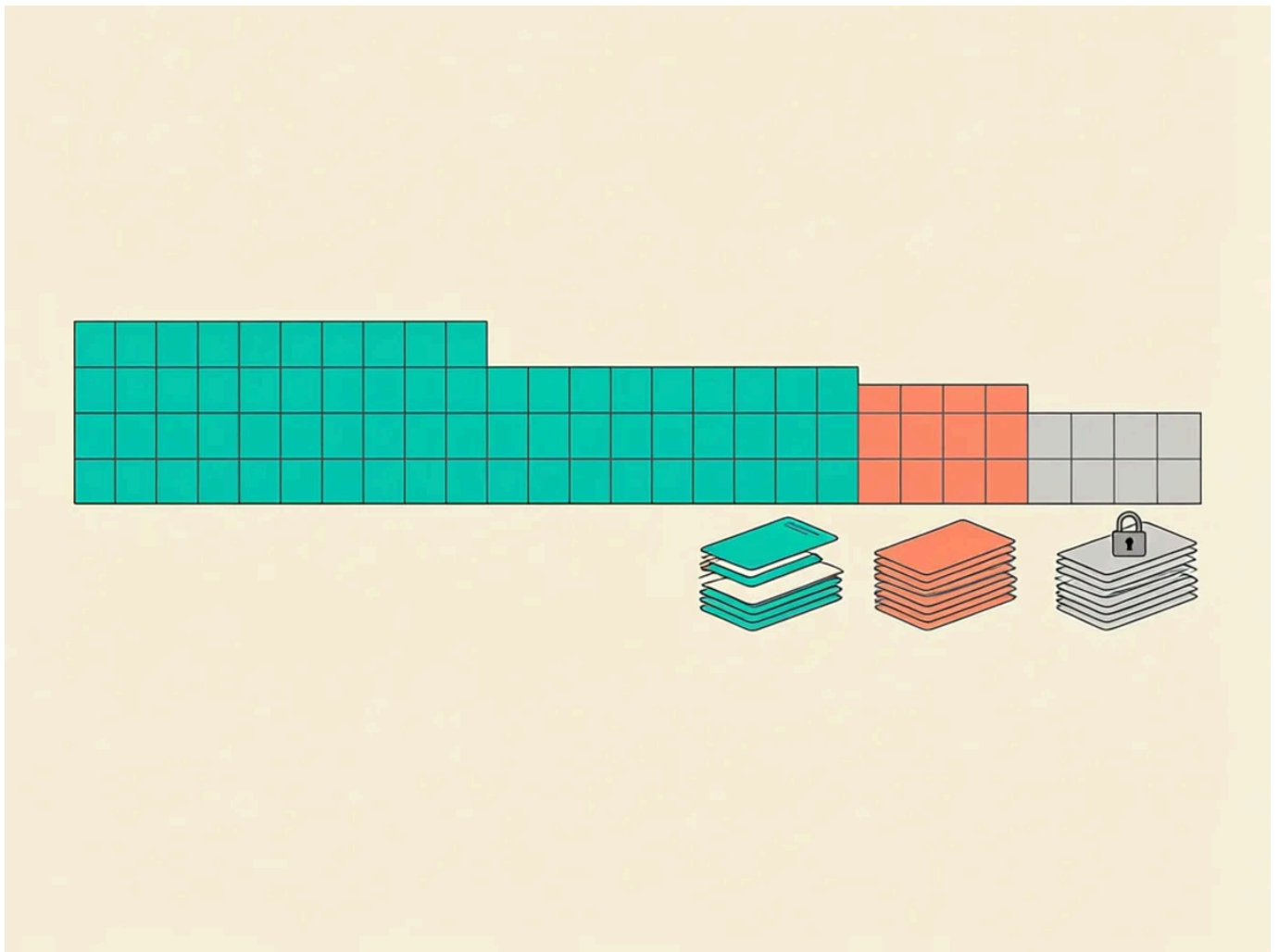
Es la mayor parte de los datos, y sirve para ajustar los parámetros del modelo. Aquí el modelo mira las respuestas y aprende. Su error suele ser bajo cuanto más flexible es el modelo, pero ese número no informa sobre el mundo exterior.

El conjunto de validación

Reservado para afinar hiperparámetros y elegir entre candidatos: probamos configuraciones, miramos el error de validación y nos quedamos con la mejor. Este conjunto guía decisiones, así que acaba filtrándose ligeramente al proceso; no sirve para la cifra final honesta.

El conjunto de prueba

Intocable hasta el final. Se usa una sola vez, para reportar el rendimiento con que nos presentamos al mundo. En cuanto lo miramos para tomar decisiones, contamina el proceso y deja de ser prueba honesta: se convierte en validación disfrazada.



La división en tres: entrenamiento (el tramo mayor), validación (para afinar) y prueba (pequeña y sellada), que solo se abre al final para un veredicto honesto.

Por qué tres y no dos

Con solo entrenamiento y prueba, cada ajuste de hiperparámetros se haría mirando la prueba, que gradualmente sobreajustaríamos. La validación absorbe esas decisiones y protege la prueba para que siga siendo un árbitro limpio.

Tamaños típicos

No hay ley universal, pero son comunes proporciones como ochenta por ciento entrenamiento, diez de validación y diez de prueba, o esquemas del noventa y cinco por ciento cuando los datos escasean. Con poco dato, reservar demasiado empobrece el entrenamiento: hay que equilibrar.

Muestreo representativo

Los tres conjuntos deben provenir de la misma distribución y, en clasificación, mantener la mezcla de clases —muestreo estratificado—. Una prueba con proporciones distintas daría una medida engañosa del comportamiento real.

El espejismo del test set único

Un solo corte de prueba da una estimación ruidosa: dos divisiones distintas cambian el número. Con pocos datos, una cifra de prueba puede ser casi azar. De ahí que a menudo se prefieran esquemas de validación cruzada más estables.

Fuga de datos

La contaminación sutil es el enemigo: estandarizar con medias calculadas sobre todos los datos, o incluir duplicados en entrenamiento y prueba a la vez, filtra información del futuro. La división debe hacerse antes de cualquier transformación que use estadísticos.

Errores frecuentes

- Reutilizar el conjunto de prueba para elegir hiperparámetros una y otra vez.
- Aplicar transformaciones ajustadas con todos los datos antes de dividir.
- Celebrar el bajo error de entrenamiento como si fuera capacidad predictiva.

Ejemplo resuelto: partir los datos de un clasificador

Se tienen diez mil clientes etiquetados y hay que medir de forma honesta. El protocolo:

1. Dividir primero: apartar prueba y validación antes de tocar nada, para no contaminarlas.

2. Estratificar: conservar en cada porción la misma proporción de clases que en el total.
3. Entrenar con la porción mayor, ajustando los parámetros del modelo.
4. Afinar con la validación: elegir configuración mirando solo ese error.
5. Cerrar la prueba con llave y abrirla una única vez para reportar la cifra final.

¿Para qué sirve en la realidad?

Toda decisión de desplegar un modelo se apoya en estimar su rendimiento futuro. Una división disciplinada evita engañarse con números inflados; una fuga de datos puede hacer fracasar en producción lo que brillaba en el laboratorio.

1. ¿Cuál es la función del conjunto de validación en la división clásica en tres partes?

- ☐ Ajustar los parámetros internos del modelo
- ☐ Afinar hiperparámetros y elegir entre candidatos
- ☐ Reportar la cifra final de rendimiento ante terceros
- ☐ Aumentar artificialmente los datos de entrada

2. Tras varias iteraciones afinando con el conjunto de prueba, este sigue siendo una medición honesta.

- ☐ Verdadero
- ☐ Falso

El precio de mirar demasiado

Cada vez que asomas la nariz al conjunto de prueba para «ver cómo va», pagas con su honestidad. La disciplina exige trabajar con validación y abrir la prueba una única vez, al final, casi con miedo. La tentación de reintentar hasta que salga bien es la trampa que destruye la validez.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Division



4. Ordena el uso de los tres conjuntos en un proyecto:

Elementos:

- repartir los datos en tres porciones al inicio
- ajustar parámetros con el conjunto mayor
- afinar hiperparámetros mirando el error de validación
- reportar una única vez el rendimiento en prueba

Dividir	. . .
Entrenar	. . .
Validar	. . .
Probar	. . .

Explained <https://explained.ai/> Explicaciones visuales sobre validación y generalización en modelos.

5. Une cada conjunto de datos con su papel en el proyecto.

5.	ajusta los parámetros internos	-----
5.	afina hiperparámetros y elige candidatos	-----
5.	intocable hasta el reporte final	-----
5.	conservar la mezcla de clases al partir	-----

Opciones: entrenamiento · validación · prueba · estratificado

Tres conjuntos

División de datos

Entrenar

la mayor parte

ajusta parámetros

Validar

afinar hiperparámetros

elegir candidatos

Probar

una sola vez

veredicto honesto

Cuidados

estratificar

evitar fuga de datos

El mapa mental interactivo es interactivo en la versión online.



Conjuntos de entrenamiento, validación y prueba https://es.wikipedia.org/wiki/Conjuntos_de_datos_de_entrenamiento,_validaci3n_y_prueba Referencia en español sobre la división de datos y sus objetivos.

6. Un compañero ajusta el modelo ochenta veces mirando el accuracy del test set y obtiene noventa y nueve por ciento. ¿Por qué desconfiarías y qué protocolo propondrías?

Referencia curricular

Grado:	5
Materia:	Fundamentos de Machine Learning · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://lumbre.lat/object/cdia-mlf-003-division-train-validation-test>

Material educativo generado desde Lumbre · lumbre.lat