

Aprendizaje automático

Evaluación de modelos de clasificación y selección de hiperparámetros

Al igual que en regresión, necesitamos de un conjunto de datos que no se haya utilizado para entrenar un modelo para poder llevar a cabo la evaluación del mismo.

Con dicho **conjunto de prueba**, calculamos algunas métricas que nos indican si el modelo generaliza correctamente.

Matriz de confusión

- Una forma de resumir los resultados de la evaluación de un modelo es a través de una **matriz de confusión**.

	Perro (Real)	No perro (Real)	Total
Perro (Predicha)	5 (TP)	1 (FP)	6
No perro (Predicha)	2 (FN)	2 (TN)	4
Total	7	3	10

- TP – Verdaderos positivos
- FP – Falsos positivos
- FN – Falsos negativos
- TN – Verdaderos negativos

Matriz de confusión

$$\text{Accuracy} = \frac{TP + TN}{\text{Total de observaciones}}$$

$$\text{Recall}_{\text{perro}} = \frac{TP}{TP + FN}$$

$$\text{Precision}_{\text{perro}} = \frac{TP}{TP + FP}$$

$$\text{Recall}_{\text{no perro}} = \frac{TN}{TN + FP}$$

$$\text{Precision}_{\text{no perro}} = \frac{TN}{TN + FN}$$

$$\text{F-measure}_i = \frac{2 * \text{Recall}_i * \text{Precision}_i}{\text{Recall}_i + \text{Precision}_i} \quad (\text{Para } i = \text{perro, no perro})$$

Matriz de confusión

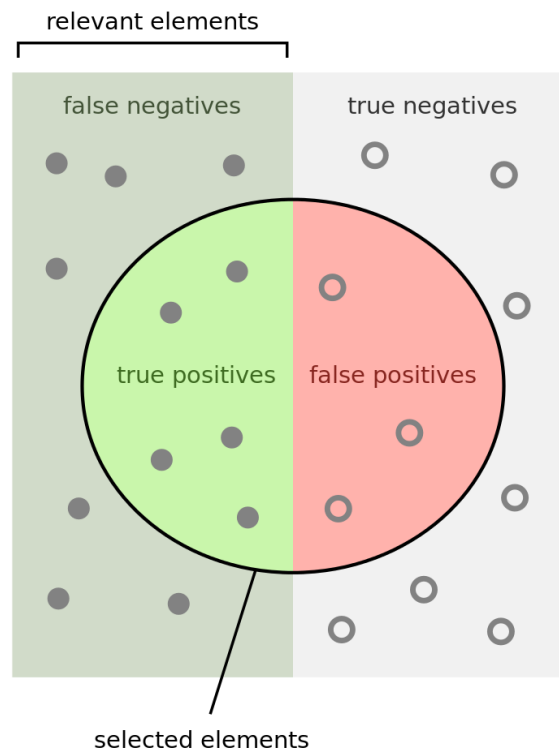
$$\text{Weighted - Accuracy} = w_P * \frac{TP}{TP + FN} + w_N * \frac{TN}{TN + FP}$$

$$w_P + w_N = 1$$

$$0 \leq w_P \leq 1$$

$$0 \leq w_N \leq 1$$

Matriz de confusión



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

[Wikipedia – Precision and recall](#)

// Matriz de confusión

	Gato (Real)	Pescado (Real)	Gallina (Real)	Total
Gato (Predicha)	4	6	3	13
Pescado (Predicha)	1	2	0	3
Gallina (Predicha)	1	2	6	9
Total	6	10	9	25

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$$

TP_i Verdaderos positivos para la clase i

FN_i Falsos negativos para la clase i

FP_i Falsos positivos para la clase i

Matriz de confusión

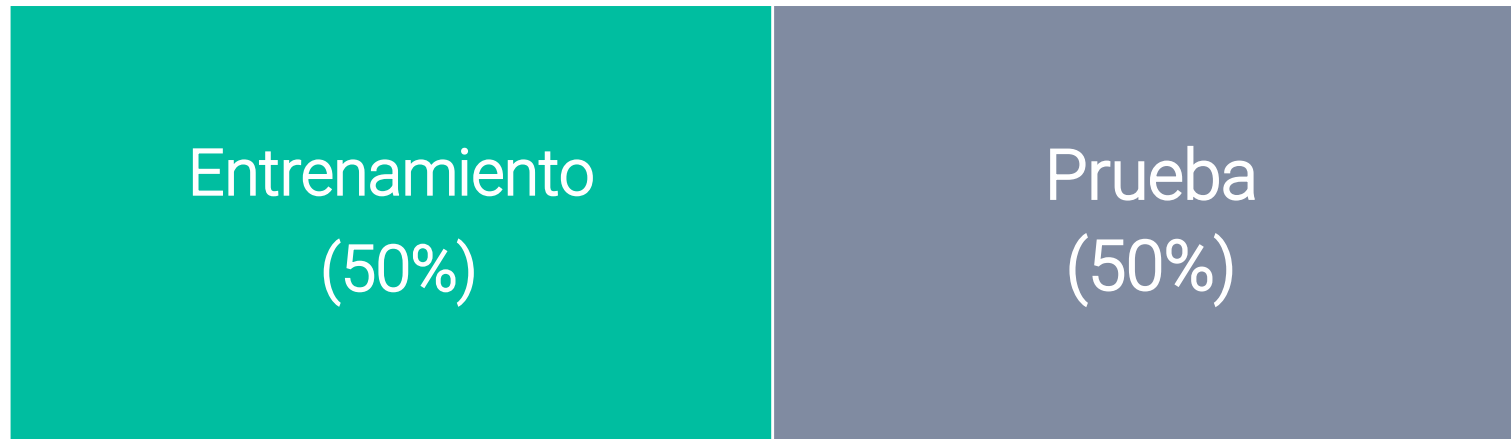
$$\text{Accuracy} = \frac{TP_1 + TP_2 + TP_3}{\text{Total de observaciones}}$$

$$\text{Weighted Accuracy} = w_1 * \frac{T_1}{T_1 + F_1} + w_2 * \frac{T_2}{T_2 + F_2} + w_3 * \frac{T_3}{T_3 + F_3}$$

$$w_1 + w_2 + w_3 = 1$$

$$0 \leq w_1 \leq 1$$

División del conjunto de datos para entrenamiento y prueba



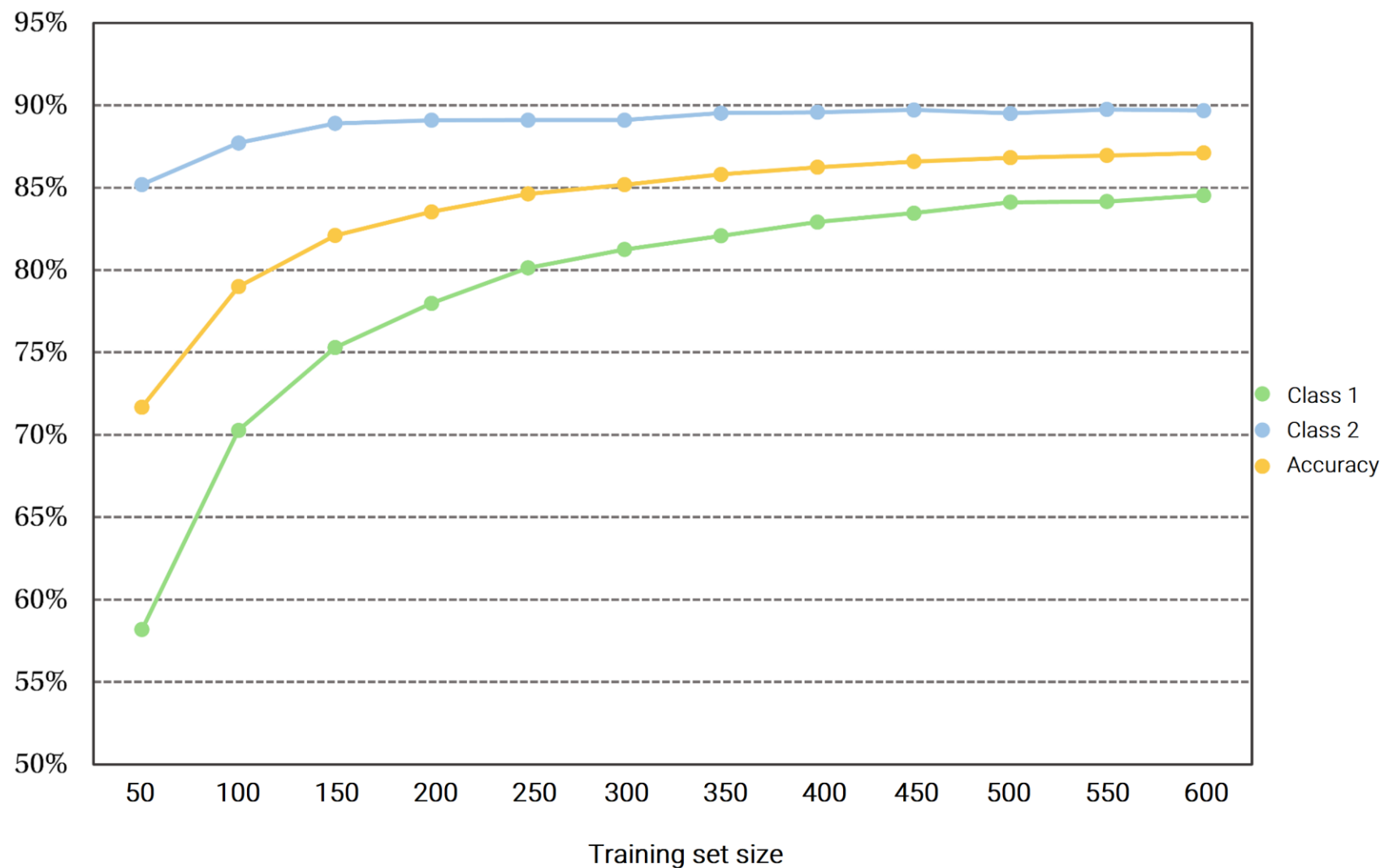
Tamaño del conjunto de entrenamiento

- Algo importante a la hora de entrenar un modelo de aprendizaje supervisado es determinar si la cantidad de muestras disponibles es suficiente para que el algoritmo de aprendizaje logre generalizar.
- La cantidad de datos necesaria para entrenar un modelo depende tanto de las **características del problema** (lineal, no lineal, etc.), la **cantidad de predictores**, y el **algoritmo de clasificación**.
- Pocas muestras producen modelos que **pobremente generalizan**, mientras que adquirir un conjunto grande de observaciones puede ser altamente **costoso**.

Para un conjunto de datos, es posible probar diferentes tamaños del conjunto de entrenamiento seleccionando aleatoriamente observaciones para generar diferentes conjuntos.

Si el rendimiento del clasificador no mejora al aumentar el número de muestras, podemos decir que tenemos las suficientes para entrenarlo. Si este no es el caso, es posible que necesitemos más muestras para mejorar nuestro modelo.

Tamaño del conjunto de entrenamiento





¿Qué pasa si mi conjunto de datos es pequeño y no lo podemos dividir en dos?

Todos los métodos de **validación cruzada** aplican para el caso de clasificación. A su vez, varios de los métodos de validación cruzada tienen una versión **estratificada**.

En los **métodos estratificados de validación cruzada**, se garantiza que en cada **partición** se mantenga la **proporción de observaciones de cada clase** que se tiene en el conjunto de datos originales.

Muestras desbalanceadas

Muestras desbalanceadas

- Es común tener una situación en la que las observaciones de una clase son mucho más que las que se tienen disponibles para otra clase (**muestra desbalanceada**). Por ejemplo, para una rara enfermedad, es más fácil encontrar gente que no la tenga que personas con dicho padecimiento.
- Debido a que muchos modelos de clasificación se entrenan teniendo en cuenta alguna medida del error de clasificación (por ejemplo, SVM), es posible que, ante una muestra desbalanceada, el clasificador tenga un **sesgo** hacia la clase con más observaciones.
- En el caso extremo, si hay 100 veces más muestras para una clase que para otra, es probable que el clasificador simplemente responda siempre hacia la clase con más muestras ya que con ello va a tener una gran exactitud.

Para entender si un clasificador está en esta situación, es importante medir la sensibilidad (recall) por clase y la exactitud ponderada para $w_P = w_N = 0.5$.

Estrategias para muestras desbalanceadas

- Submuestrear la clase (o las clases) con más muestras.
 - Se selecciona un subconjunto de muestras de las clases con más observaciones de tal forma que el conjunto de entrenamiento tenga la misma cantidad de muestras para todas las clases.
- Sobremuestrear la clase (o las clases) con menos muestras.
 - Se selecciona un subconjunto de muestras de las clases con menos observaciones utilizando muestreo con remplazo para equilibrar la cantidad de muestras.

Estrategias para muestras desbalanceadas

- Modificar la función de costo del modelo o adaptar el algoritmo de aprendizaje.
 - Modificar el algoritmo de aprendizaje para balancear la importancia de las clases de acuerdo al número de muestras disponibles.
 - Clasificadores como SVM, regresión logística, redes neuronales y árboles de decisión permiten de manera natural hacer el ajuste para muestras desbalanceadas.
 - Modelos como LDA, QDA, etc., permiten el ajuste modificando lo que se conoce como probabilidades a priori de las clases.

El submuestreo y sobremuestreo se realizan únicamente sobre el conjunto de entrenamiento que se pasa al algoritmo de optimización, incluyendo en validación cruzada.

No se debe considerar que la muestra submuestreada o sobremuestreada es un nuevo conjunto de datos para llevar a cabo todo el análisis.

Selección de hiperparámetros

Hiperparámetros

- Algunos modelos tienen parámetros que no pueden ser estimados en el proceso de entrenamiento.
- Por ejemplo, la k de los modelos K-NN, el parámetro γ de RB-SVM, el parámetro C de SVM, el parámetro λ del clasificador logístico, el número de capas de una red neuronal, el parámetro λ de un modelo de regresión lineal regularizado, etc.
- También se desconoce otros elementos importantes como el número óptimo de características a escoger en los métodos de selección de características.
- A estos parámetros se les conoce como hiperparámetros.

Para calcular la configuración óptima de hiperparámetros, es necesario evaluar varias configuraciones, y tomar aquella que de los mejores resultados.

Problemas en la selección de hiperparámetros

- Para cada configuración de hiperparámetros se podría utilizar el conjunto de entrenamiento para hacer el ajuste del modelo y evaluar con el conjunto de prueba. De aquí se seleccionaría la mejor configuración.
- Sin embargo, esto **sobreestimaría** artificialmente el rendimiento del modelo.
- Los hiperparámetros son otro elemento a aprender en la generación de un modelo, por lo que no se debe utilizar el conjunto de prueba para hacer la optimización de los mismos.
- Es necesario otro conjunto de datos llamado **conjunto de validación**, con el cual se evalúan las diferentes configuraciones del clasificador.

Conjuntos de datos en aprendizaje supervisado

Conjunto de entrenamiento

Para entrenar el modelo y encontrar sus parámetros.

Conjunto de validación

Para encontrar los hiperparámetros del modelo.

Conjunto de prueba

Para estimar el rendimiento del modelo y su capacidad de generalización.

Conjuntos de datos en aprendizaje supervisado





¿Qué pasa si mi conjunto de datos es pequeño y no lo podemos dividir en tres?

Se requiere **validación cruzada anidada** para este caso.

Esto se debe a que los hiperparámetros óptimos se encuentran con validación cruzada dentro de otra validación cruzada en el que se lleva a cabo la evaluación del modelo con los hiperparámetros óptimos.

Selección de hiperparámetros con validación cruzada anidada

- Por cada partición aleatoria
 - Por cada configuración de hiperparámetros
 - Por cada subpartición del subconjunto de entrenamiento
 - Entrenar con el subsubconjunto de entrenamiento.
 - Evaluar con el subsubconjunto de validación.
 - Seleccionar la mejor configuración de hiperparámetros.
 - Entrenar con el subconjunto de entrenamiento y los hiperparámetros óptimos.
 - Evaluar con el subconjunto de prueba.
- Reportar el rendimiento promedio de todas las particiones.

Selección de hiperparámetros y de características con validación cruzada anidada

- Por cada partición aleatoria
 - Por cada configuración de hiperparámetros
 - Por cada subpartición del subconjunto de entrenamiento
 - Seleccionar características con el subsubconjunto de entrenamiento.
 - Entrenar con el subsubconjunto de entrenamiento y las características encontradas.
 - Evaluar con el subsubconjunto de validación.
 - Seleccionar la mejor configuración de hiperparámetros.
 - Seleccionar características con el subconjunto de entrenamiento.
 - Entrenar con el subconjunto de entrenamiento, y los hiperparámetros óptimos y las características encontradas.
 - Evaluar con el subconjunto de prueba.
- Reportar el rendimiento promedio de todas las particiones.

Si no se requiere hacer la evaluación del modelo, podemos omitir el ciclo más exterior de la validación cruzada.

Selección de hiperparámetros con validación cruzada sin evaluación del modelo

- Por cada configuración de hiperparámetros
 - Por cada partición aleatoria
 - Entrenar con el subconjunto de entrenamiento.
 - Evaluar con el subconjunto de validación.
- Seleccionar la mejor configuración de hiperparámetros.
- Entrenar con todos los datos y los hiperparámetros óptimos.

Selección de hiperparámetros y de características con validación cruzada sin evaluación del modelo

- Por cada configuración de hiperparámetros
- Por cada partición aleatoria
 - Seleccionar características con el subconjunto de entrenamiento.
 - Entrenar con el subconjunto de entrenamiento y las características encontradas.
 - Evaluar con el subconjunto de validación.
- Seleccionar la mejor configuración de hiperparámetros.
- Seleccionar características con todos los datos.
- Entrenar con todos los datos, los hiperparámetros óptimos y las características encontradas.

Ya que se haya evaluado el modelo correctamente, se procede al ajuste del modelo para producción.

Flujos de trabajo para la generación de un modelo de aprendizaje automático

Si se tienen muchísimos datos

Dividir los datos en conjuntos de entrenamiento y prueba.



Ajustar el modelo con el conjunto de entrenamiento y evaluar el modelo con el conjunto de prueba.



Usar el modelo anterior para producción. Otra opción es entrenar el modelo de nuevo con todos los datos.

Si se tienen muchísimos datos y se requiere selección de características

Dividir los datos en conjuntos de entrenamiento y prueba.

Seleccionar características para el conjunto de entrenamiento.

Con el conjunto de características encontrados, ajustar el modelo con el conjunto de entrenamiento y evaluar el modelo con el conjunto de prueba.

Usar el modelo anterior para producción. Otra opción es entrenar el modelo de nuevo con todos los datos, pero se debe realizar la selección de características de nuevo.

Si se tienen muchísimos datos y se requiere selección de características (rentrenamiento para producción)

Seleccionar características con todos los datos.



Con las características seleccionadas, entrenar el modelo con todos los datos para producción.

Si se tienen muchísimos datos y se requieren hiperparámetros

Dividir los datos en conjuntos de entrenamiento, validación y prueba



```
graph TD; A[Dividir los datos en conjuntos de entrenamiento, validación y prueba] --> B[Buscar el conjunto de hiperparámetros óptimo utilizando el conjunto de entrenamiento para ajustar los modelos y el de validación para evaluar las diferentes configuraciones de hiperparámetros.]; B --> C[Con el conjunto óptimo de hiperparámetros, ajustar el modelo con el conjunto de entrenamiento y evaluar el modelo con el conjunto de prueba.]; C --> D[Usar el modelo anterior para producción. Otra opción es entrenar el modelo de nuevo con todos los datos, pero en este caso, se deben encontrar de nuevo los hiperparámetros del modelo.];
```


Buscar el conjunto de hiperparámetros óptimo utilizando el conjunto de entrenamiento para ajustar los modelos y el de validación para evaluar las diferentes configuraciones de hiperparámetros.

Con el conjunto óptimo de hiperparámetros, ajustar el modelo con el conjunto de entrenamiento y evaluar el modelo con el conjunto de prueba.

Usar el modelo anterior para producción. Otra opción es entrenar el modelo de nuevo con todos los datos, pero en este caso, se deben encontrar de nuevo los hiperparámetros del modelo.

Si se tienen muchísimos datos y se requieren hiperparámetros (rentrenamiento para producción)

Dividir los datos en entrenamiento y validación.



Buscar los hiperparámetros óptimos utilizando el conjunto de entrenamiento para ajustar el modelo y el conjunto de validación para evaluar las diferentes configuraciones de hiperparámetros.



Con los hiperparámetros encontrados, entrenar el modelo con todos los datos para producción.

Si se tienen muchísimos datos y se requiere selección de características e hiperparámetros

Dividir los datos en conjuntos de entrenamiento, validación y prueba

Buscar el conjunto de hiperparámetros óptimo utilizando el conjunto de entrenamiento para seleccionar características y ajustar los modelos, y el de validación para evaluar las diferentes configuraciones de hiperparámetros

Con el conjunto óptimo de hiperparámetros, seleccionar características con el conjunto de entrenamiento.

Con el conjunto óptimo de hiperparámetros y las características encontradas, ajustar el modelo con el conjunto de entrenamiento y evaluar el modelo con el conjunto de prueba.

Usar el modelo anterior para producción. Otra opción es entrenar el modelo de nuevo con todos los datos, pero en este caso, se deben encontrar de nuevo los hiperparámetros del modelo y seleccionar características

Si se tienen muchísimos datos y se requiere selección de características e hiperparámetros (rentrenamiento para producción)

Dividir los datos en entrenamiento y validación.



```
graph TD; A[Dividir los datos en entrenamiento y validación.] --> B[Buscar los hiperparámetros óptimos utilizando el conjunto de entrenamiento para seleccionar características y ajustar el modelo, y el conjunto de validación para evaluar las diferentes configuraciones de hiperparámetros.]; B --> C[Con los hiperparámetros óptimos, seleccionar características con todos los datos.]; C --> D[Con los hiperparámetros encontrados y las características encontradas, entrenar el modelo con todos los datos para producción.];
```

Buscar los hiperparámetros óptimos utilizando el conjunto de entrenamiento para seleccionar características y ajustar el modelo, y el conjunto de validación para evaluar las diferentes configuraciones de hiperparámetros.

Con los hiperparámetros óptimos, seleccionar características con todos los datos.

Con los hiperparámetros encontrados y las características encontradas, entrenar el modelo con todos los datos para producción.

Cuando hay muchos datos, en la practica, no se rentrena el modelo con todos los datos ya que al tener un numero grande de observaciones, no se espera que el modelo cambie mucho respecto a la versión de cuando se entrenó y validó.

Si se tienen pocos datos

Evaluar el modelo con validación cruzada.




Entrenar el modelo con todos los datos para producción.

Si se tienen pocos datos y se requiere selección de características

Evaluar el modelo con validación cruzada.



Seleccionar características con todos los datos.



Con las características seleccionadas, entrenar el modelo con todos los datos para producción.

Si se tienen pocos datos y se requieren hiperparámetros

Evaluar el modelo con validación cruzada anidada.



Buscar el conjunto de hiperparámetros óptimo utilizando validación cruzada con todos los datos.



Con los hiperparámetros óptimos, entrenar el modelo con todos los datos para producción.

Si se tienen pocos datos y se requiere selección de características e hiperparámetros

Evaluar el modelo con validación cruzada anidada.

Buscar el conjunto de hiperparámetros óptimo utilizando validación cruzada con todos los datos y haciendo selección de características en cada iteración de la validación.

Con el conjunto óptimo de hiperparámetros, seleccionar características con todos los datos.

Con el conjunto óptimo de hiperparámetros y las características encontradas, ajustar el modelo con todos los datos para producción.

Una alternativa a buscar de nuevo hiperparámetros y seleccionar características después de la validación cruzada anidada es utilizar los resultados obtenidos en dichos procesos para pasar directo al entrenamiento del modelo.

Bibliografía

- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2023). *An introduction to statistical learning: with applications in Python* (2da ed.). Springer.
 - Capítulo 4
 - Capítulo 5
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2da ed.). Springer.
 - Capítulo 4
 - Capítulo 7