

ACÀMICA

TEMA DEL DÍA

Sesgo y varianza

Si bien al principio puede parecer abstracto, estos son los *enemigos* a los cuáles nos enfrentamos.

Bueno, a veces serán aliados. Pronto veremos cómo.



Agenda

Daily

Explicación: Trade-off entre Sesgo y Varianza

Break

Hands-on training

Segundo Proyecto

Cierre



Daily



Daily



Sincronizando...

Bitácora



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Challenge



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

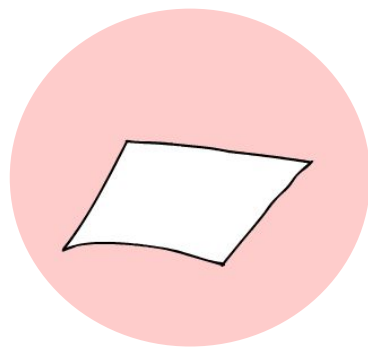
Repaso

¿Qué es *aprender* en
Machine Learning?



Ahora que ya conocemos algunos modelos, tratemos de darles un marco más general.¹

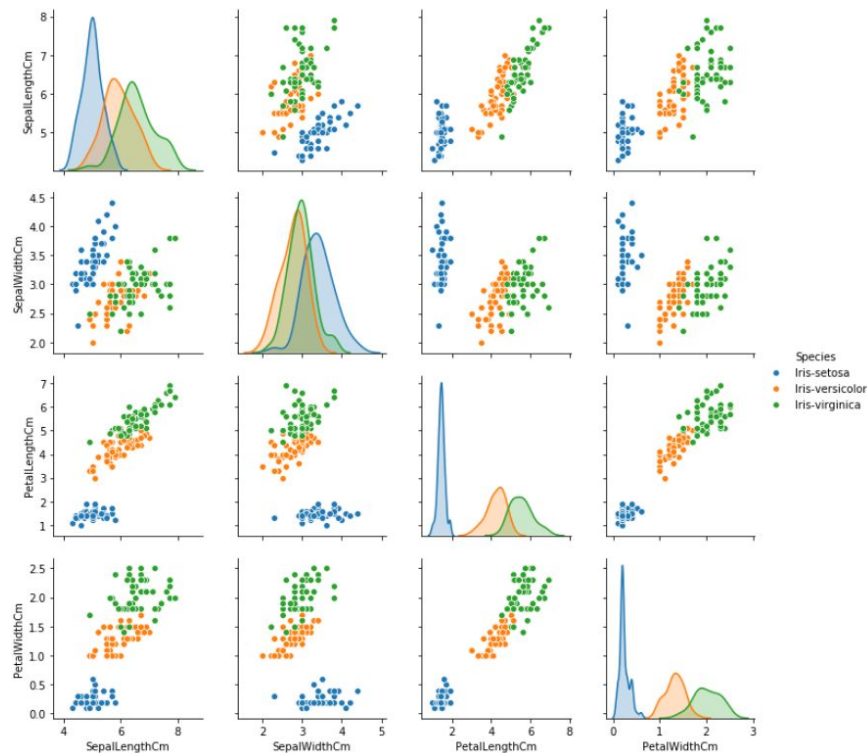
¹ Algunas ideas y textos para esta clase fueron tomadas de la materia Aprendizaje Automático, dictada por Agustín Gravano y Pablo Brusco en FCEN, UBA.



Aprendizaje **inductivo**

En el Iris Dataset, vemos que algunas características - atributos - de las flores sirven para diferenciar las especies.

Si consideramos sólo dos atributos, podemos ver que existen regiones en el plano que, aproximadamente, corresponden a cada especie.



Aprendizaje **inductivo**

La idea subyacente es que podemos aprender a diferenciar cada especie a partir de medir algunas de sus características¹.

Más general, podemos aprender conceptos a partir de un conjunto de ejemplos y sus características.

¹ Desde otro punto de vista, si esas características que medimos nos sirven para diferenciar, tal vez sean buenas características para definir esas especies.

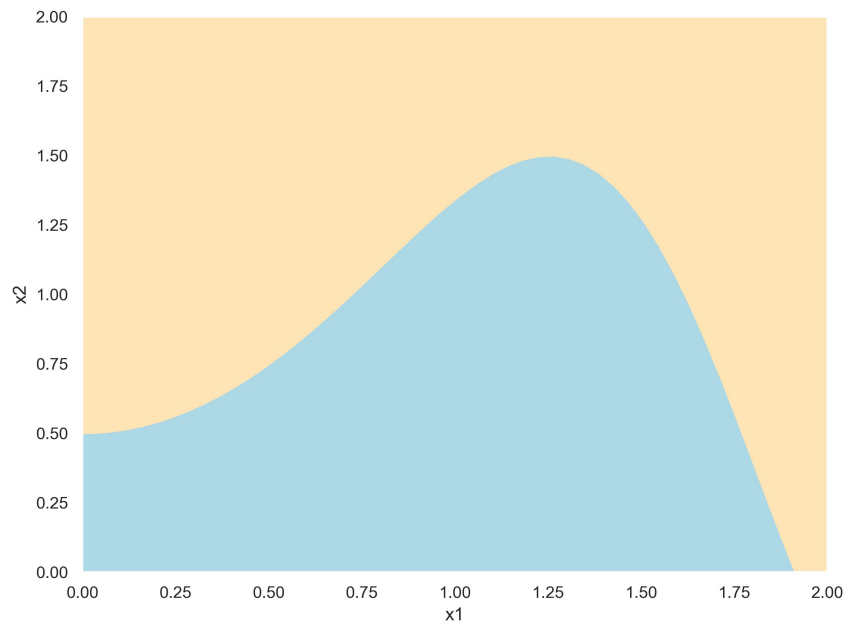
Aprendizaje **inductivo**

Usando esas características (o propiedades) que medimos, asumimos que existe una frontera que separa las clases que queremos aprender.

Aprendizaje **inductivo**

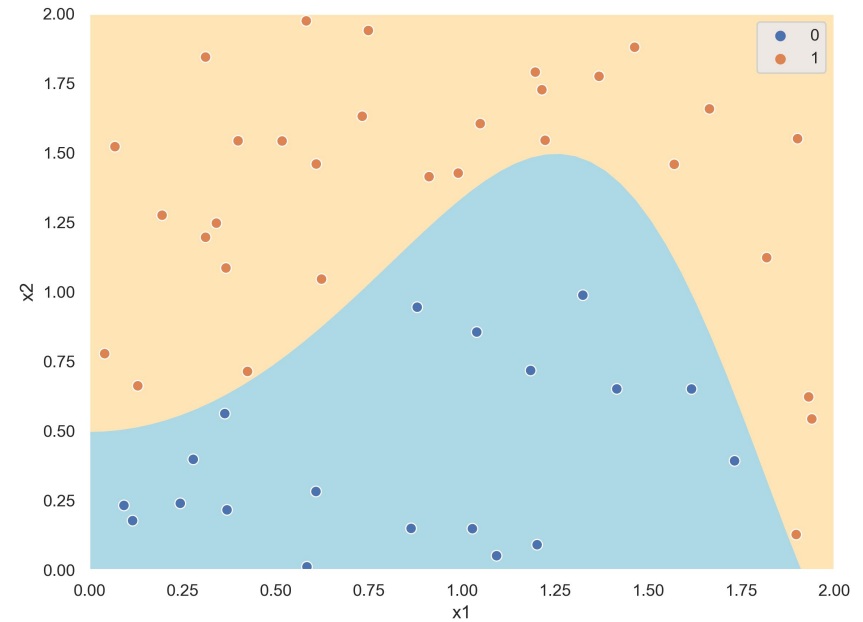
Veamos un ejemplo de juguete.

Queremos aprender a qué regiones corresponde el color amarillo y a qué regiones corresponde el color celeste.
Es decir, **$f(x_1, x_2) \rightarrow \text{color}$**



Aprendizaje inductivo

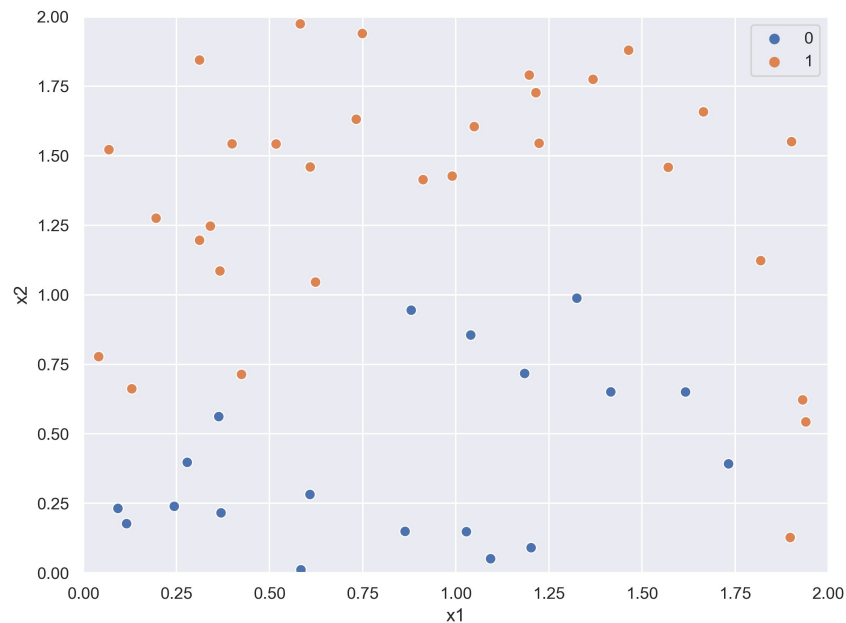
Pero no tenemos las regiones pintadas, sino muestras.



Aprendizaje **inductivo**

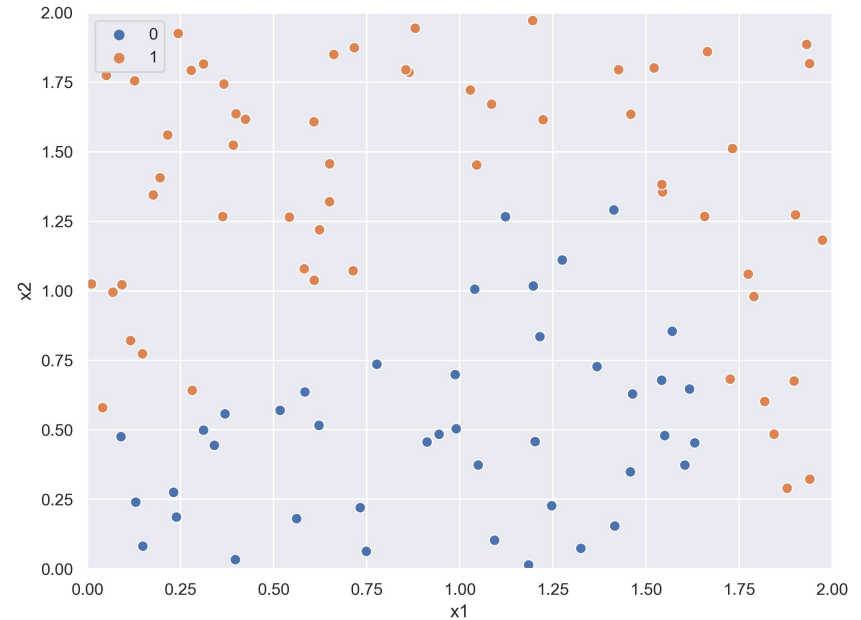
Pero no tenemos las regiones pintadas, sino muestras.

Muestras sobre un fondo sin color.



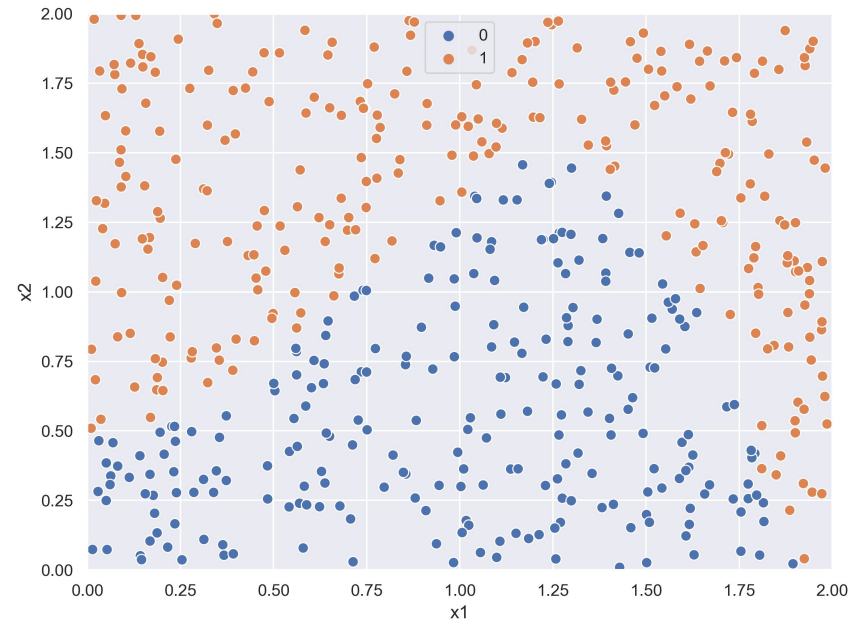
Aprendizaje inductivo

Cuantas más muestras,
probablemente sea más fácil la tarea.



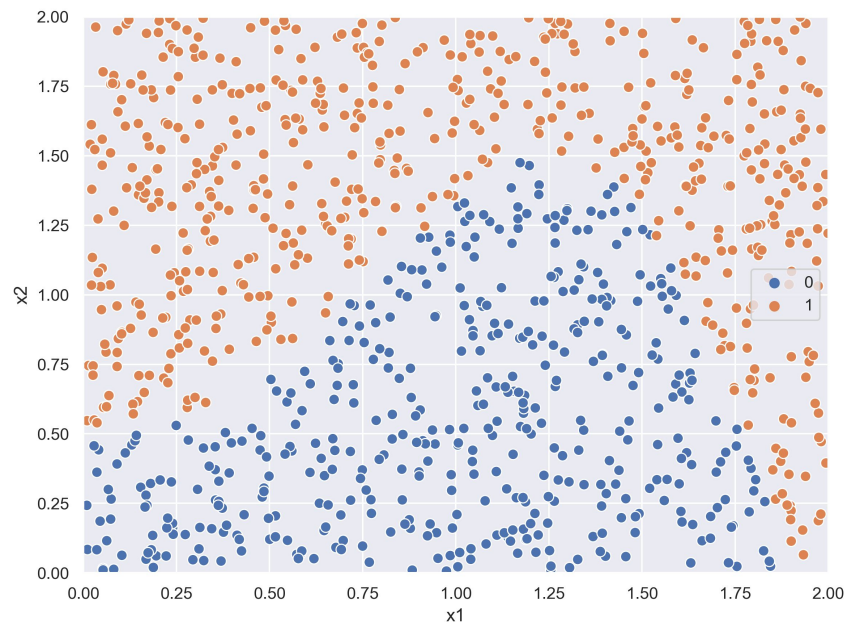
Aprendizaje inductivo

Cuantas más muestras,
probablemente sea más fácil la tarea.



Aprendizaje **inductivo**

Cuantas más muestras,
probablemente sea más fácil la tarea.



El aprendizaje inductivo consiste en

construir un modelo general (color del fondo)
a partir de información específica (instancias).

Principio de Aprendizaje Inductivo

Cualquier modelo que aproxime bien a una función objetivo sobre un conjunto suficientemente grande de datos también aproximará bien a la función objetivo sobre datos no observados.

Aprendizaje **supervisado**

Dada una función objetivo **f** desconocida, queremos aproximarla mediante un modelo, que se suele notar **f^1** .

Entrenar un modelo

→
consiste en

ajustar sus parámetros (encontrar valores óptimos) dado un conjunto de datos.

Los algoritmos de aprendizaje automático son procedimientos para entrenar modelos a partir de un conjunto de datos

¹El sombrerito va a arriba de la **f** pero no encontramos cómo escribirlo.

Aprendizaje **supervisado**



¡Modelo y algoritmo no son lo mismo!

(aunque a veces los vamos a usar como sinónimos en un abuso de lenguaje).

Sesgo inductivo

Pero hay muchas formas para construir un modelo. Por ejemplo, árboles de decisión y vecinos más cercanos aprenden de los datos de formas distintas.

Sesgo inductivo

Pero hay muchas formas para construir un modelo. Por ejemplo, árboles de decisión y vecinos más cercanos aprenden de los datos de formas distintas.

El **sesgo inductivo** de un algoritmo de aprendizaje es el conjunto de afirmaciones que el algoritmo utiliza para construir un modelo

Sesgo inductivo

Pero hay muchas formas para construir un modelo. Por ejemplo, árboles de decisión y vecinos más cercanos aprenden de los datos de formas distintas.

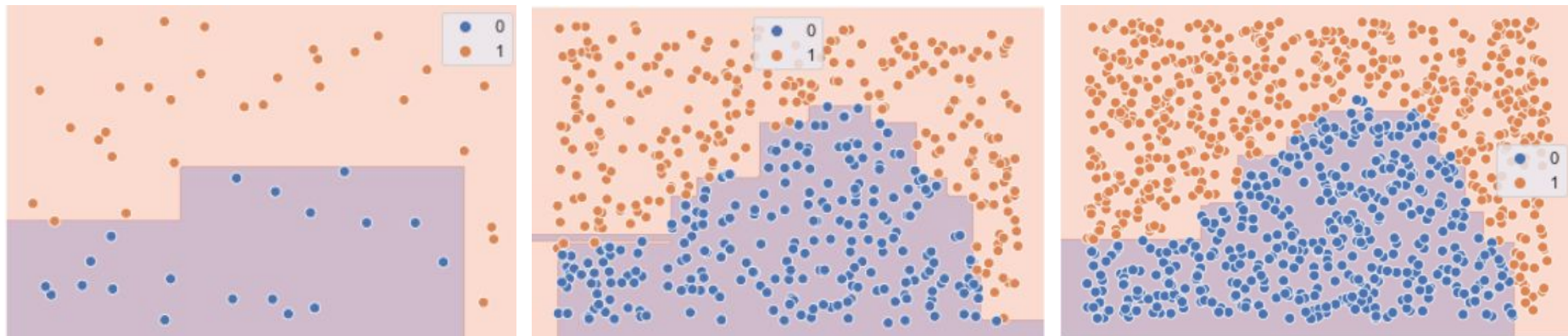
El **sesgo inductivo** de un algoritmo de aprendizaje es el conjunto de afirmaciones que el algoritmo utiliza para construir un modelo

Incluye:

- Forma de las hipótesis (número y tipo de parámetros)
- Características del funcionamiento del algoritmo (cómo recorre el espacio de hipótesis para elegir un único modelo).

Sesgo inductivo

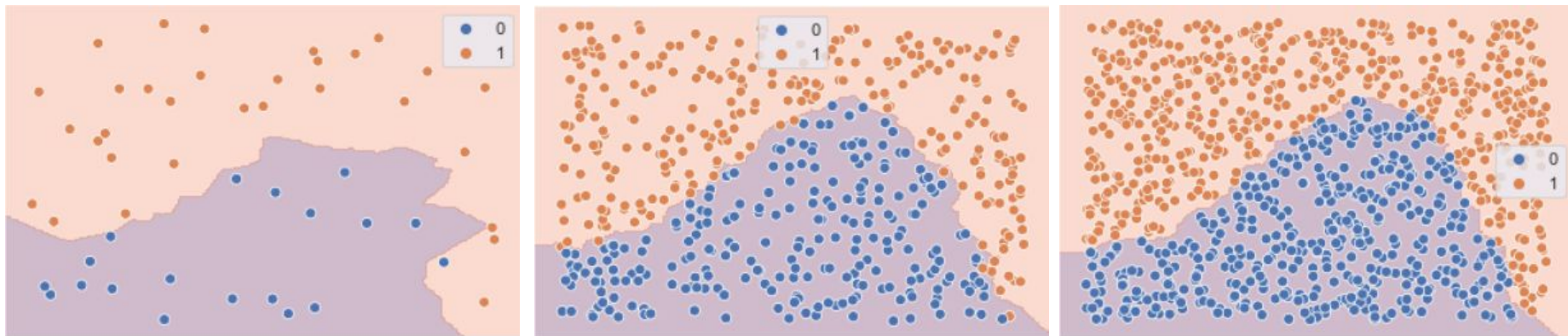
Veamos algunos ejemplos. ¿Qué ocurre si ajustamos un árbol de decisión al ejemplo que inventamos?



El espacio de hipótesis de un árbol de decisión sólo permite fronteras formadas por rectas horizontales y verticales. Podemos modificar estas fronteras variando la profundidad del árbol o la cantidad de instancias.

Sesgo **inductivo**

¿Y si ajustamos un modelo de vecinos más cercanos?



El espacio de hipótesis de vecinos más cercanos permite fronteras más versátiles que los árboles (¡esto no significa que sea un mejor modelo!). Y podemos modificar estas fronteras variando la cantidad de vecinos (k) o la cantidad de instancias.

Navaja de Ockham

En igualdad de condiciones (por ejemplo, igual desempeño), elegir la explicación (modelo) más simple.

¿Por qué? Simplemente porque esperamos que generalice mejor.

Es un principio metodológico.

Trade-off entre sesgo y varianza



Trade-off entre sesgo y varianza

No se puede construir un modelo “**perfecto**” ya que nunca pueden estar libre de errores.

Trade-off entre sesgo y varianza

No se puede construir un modelo “**perfecto**” ya que nunca pueden estar libre de errores.

Comprender cómo son - y cómo influyen - las diferentes fuentes de errores nos ayudará a mejorar el proceso de ajuste de datos, lo que resultará en mejores modelos y, adicionalmente, también evitará el sobreajuste (overfitting) y falta de ajuste (underfitting).

Error de predicción

El error de predicción para cualquier algoritmo de Machine Learning se puede dividir en tres partes:

Error irreducible (ruido)	Error de bias (sesgo)	Error de varianza
----------------------------------	------------------------------	--------------------------

Error de predicción • Error irreducible

Error irreducible (ruido)

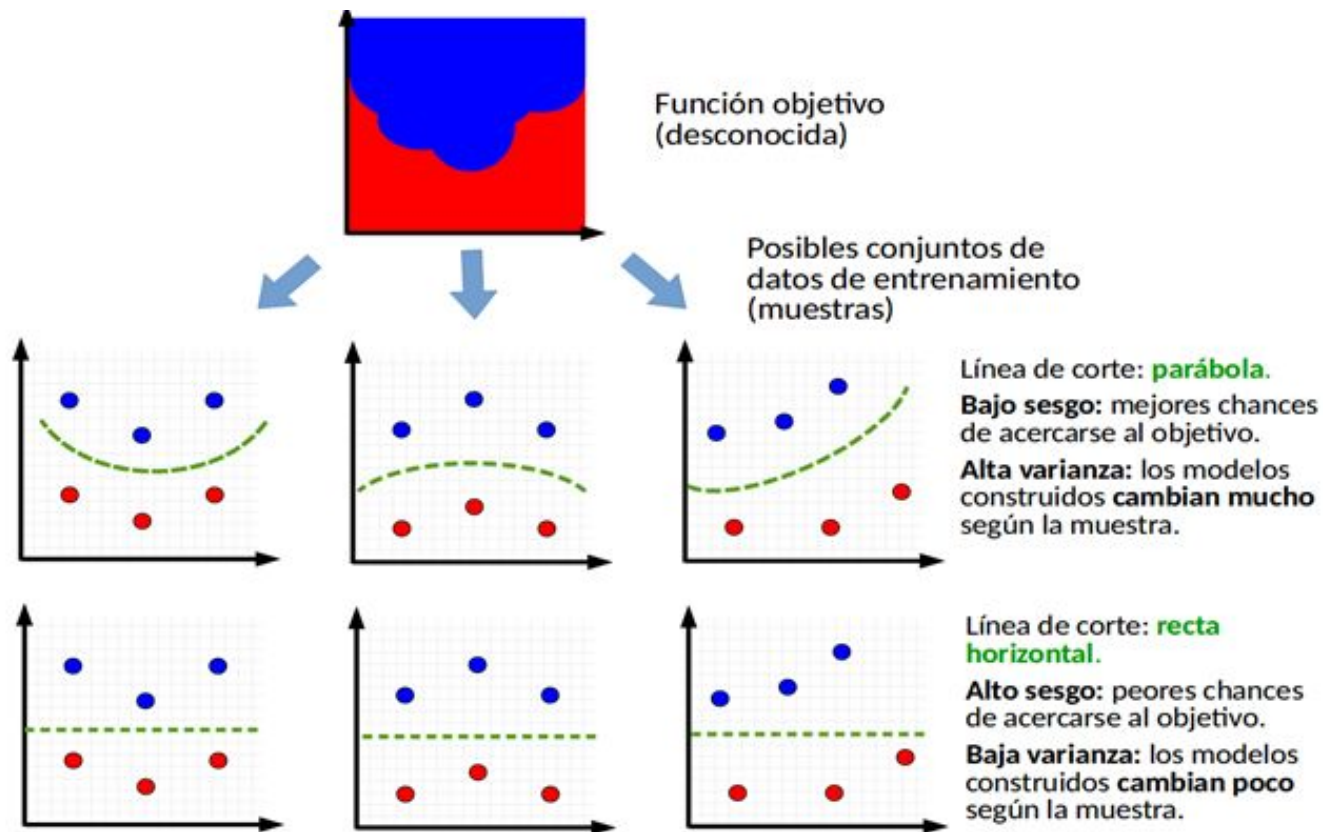
- No se puede reducir
- Se lo conoce como Ruido
- Proviene de factores como:
 1. Variables desconocidas que influyen en el mapeo de las variables de entrada a la variable de salida.
 2. Un conjunto de características incompletos
 3. Un problema mal enmarcado

No importa cuán bien hagamos nuestro modelo, nuestros datos tendrán cierta cantidad de ruido o un error irreducible que no se puede eliminar.

Errores "reducibles"

En cambio, los errores de sesgo y varianza se pueden reducir porque se derivan de la elección del algoritmo de entrenamiento y del modelo.





Error de predicción • Error de sesgo

Error de bias (sesgo)

- El sesgo es el error que introducimos al intentar explicar un problema del mundo real al que le correspondería un modelo complejo con un modelo bastante más simple.
- En general, modelos más flexibles y más complejos, implican menos sesgo.

El sesgo, entonces, no es consecuencia de nuestros datos, sino del modelo elegido.

Error de predicción • Error de sesgo

Bajo BIAS

Sugiere menos suposiciones sobre la forma de la función objetivo

Árboles de decisión, k-vecinos más cercanos y máquinas de vectores de soporte

Alto BIAS

Sugiere más suposiciones sobre la forma de la función objetivo

Regresión lineal, análisis discriminante lineal y regresión logística

Error de predicción • Error de varianza

Error de varianza

- La varianza es la cantidad en la que cambiaría mi predicción si hubiéramos entrenado nuestro modelo con un conjunto de datos diferente.
- Por ejemplo, un modelo que se ajuste mucho a unos datos sufrirá una varianza considerable al cambiar dichos datos, y viceversa.

En general, modelos más flexibles y complejos implican una mayor varianza.

Error de predicción • Error de varianza

Varianza baja

Sugiere pequeños cambios en la estimación de la función objetivo con cambios en el conjunto de datos de capacitación

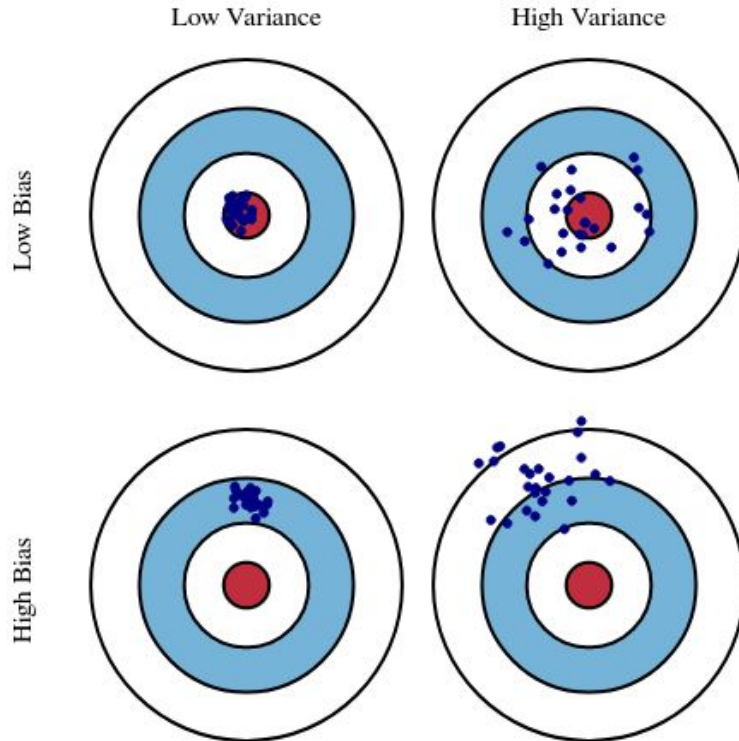
Regresión lineal, análisis discriminante lineal y regresión logística

Alta varianza

Sugiere grandes cambios en la estimación de la función objetivo con cambios en el conjunto de datos de capacitación

Árboles de decisión, k-vecinos más cercanos y máquinas de vectores de soporte

Trade-off entre sesgo y varianza



En esta gráfica se observa claramente el efecto del sesgo y la varianza sobre la predicción de un modelo, donde los puntos sobre la diana son diferentes ejecuciones de nuestro modelo.

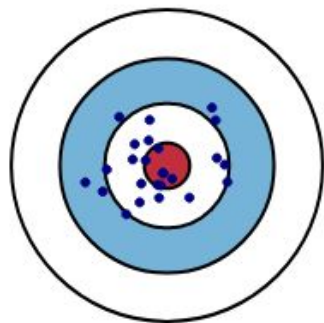
BAJA VARIANZA - ALTO BIAS



Los algoritmos de baja varianza (alto bias) tienden a ser **menos complejos**, con una estructura subyacente simple o rígida.

Entrenan modelos que son consistentes, pero inexactos en promedio. Estos incluyen algoritmos paramétricos o lineales, como la regresión lineal y Naive Bayes.

ALTA VARIANZA - BAJO BIAS



Los algoritmos de bajo bias (alta varianza) tienden a ser **más complejos**, con una estructura subyacente flexible.

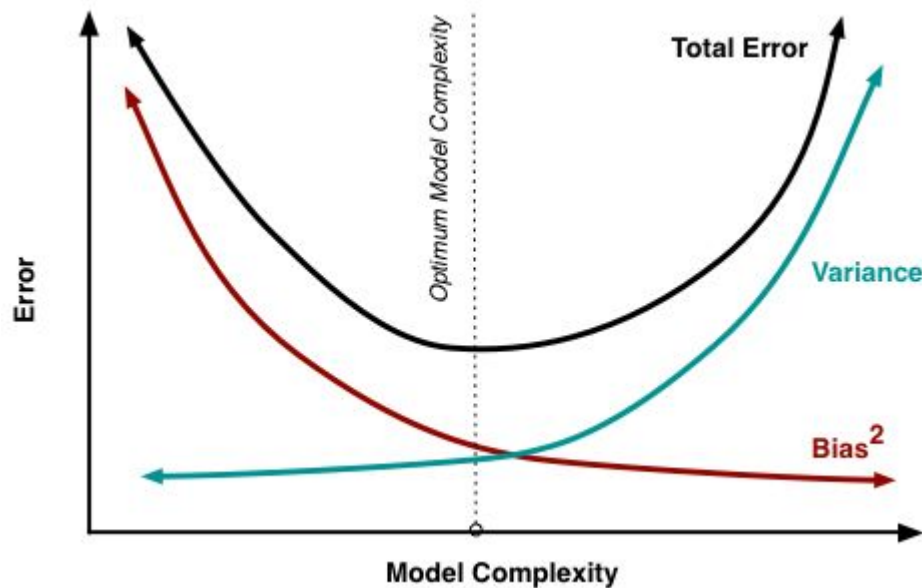
Entrenan modelos que son exactos en promedio pero inconsistentes. Estos incluyen algoritmos no lineales o no paramétricos, como árboles de decisión y k-vecinos más cercanos.

Entonces...

Cuanto más flexible sea nuestro modelo, la varianza aumentará, y el sesgo disminuirá.

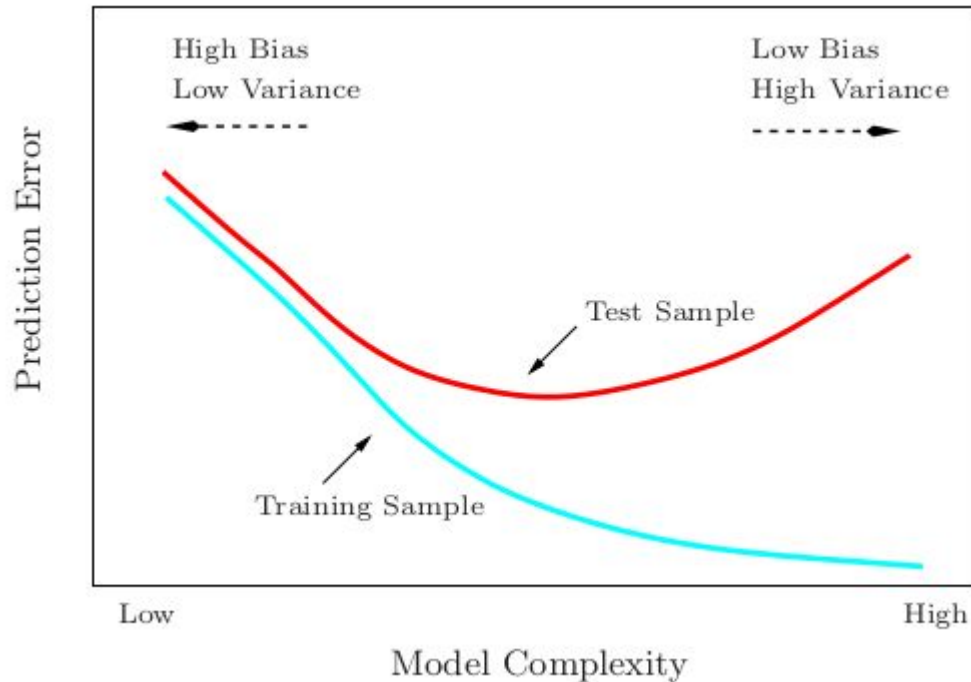
Por lo tanto, debemos prestar atención a estas características, ya que pueden introducir mucho ruido en nuestro modelo y, por lo tanto, hacerlo menos exacto.

Con un análisis de ambos podemos evaluar el rendimiento de nuestro modelo.



[Fuente](#)

Comprender el bias y la varianza es fundamental para comprender el comportamiento de los modelos de predicción, pero en general **lo que realmente importa es el error general, no la descomposición específica**. El punto ideal para cualquier modelo es el nivel de complejidad en el que el aumento en el bias es equivalente a la reducción en la varianza.



[Fuente](#)

En el gráfico, si nos movemos de izquierda a derecha:

- Aumenta la complejidad de nuestro modelo
- Baja el sesgo y aumenta la varianza.
- Hasta que llega un momento en el que el error en los datos de test empieza a aumentar mientras que el de entrenamiento sigue disminuyendo. Ese punto mínimo de error en los datos de test nos indica el nivel de complejidad óptimo para nuestro modelo.

Resumen

Modelo sesgado: No logra capturar la forma de los datos. En general, tiene desempeño muy similar en el set de entrenamiento y de validación. Asociado al underfitting.

Modelo con mucha varianza: Demasiado ajustado a los datos . Tiene desempeño muy bueno en el set de entrenamiento y malo en el de validación. Asociado al overfitting.

¿Cómo diagnosticamos sesgo y varianza?

Curva de validación/complejidad: Score en función de la complejidad. Sirve para ver regiones de baja complejidad (sesgo, underfitting) y demasiada complejidad (alta varianza, overfitting)

Curva de aprendizaje: Score en función de la cantidad de datos. Sirve para ver, dado un modelo fijo, cómo reacciona a distintos tamaño del dataset. En particular, útil para diagnosticar alta varianza o modelo muy complejo (dado el tamaño de nuestro dataset).

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and warm, creating a cozy atmosphere.

¡BREAK!

Ph. Credit: Drew Coffmann



Hands-on training



Trabajar en este notebook es opcional.
Sino, puedes trabajar en el proyecto.

`DS_Encuentro_26_Sesgo_Varianza.ipynb`



Proyecto 2



Para realizar el proyecto debes usar como referencia el contenido de las bitácoras, de los notebooks trabajados y las presentaciones vistas en clase.

También te será de mucha utilidad consultar la documentación de las librerías y de la comunidad (Stack Overflow y Google).

Encontrá tu próxima
casa en Properati



PROPERATI
EL MEJOR CAMINO A TU PRÓXIMA CASA

Q Buscar

¡Bienvenido/a a tu
~~primer~~ segundo
flujo de trabajo de
Data Science!

¡Es hora de mejorar
lo que hiciste!

Flujo de trabajo

PARTE 1 Transformación de Datos

PARTE 2 Modelos Avanzados

PARTE 3 Interpretación de Modelos

Recursos



Recursos

De Tradeoff Sesgo-Varianza:

- [WTF is the Bias-Variance Tradeoff? \(Infographic\)](#)
- [Bias and Variance – Cut Through the Noise](#)



Para la próxima

- Termina el notebook de hoy.
- Lee la bitácora 27 y carga las dudas que tengas al Trello.

En el encuentro que viene uno/a de ustedes será seleccionado/a para mostrar cómo resolvió el challenge de la bitácora. De esta manera, ¡aprendemos todos/as de (y con) todas/as, así que vengan preparados/as.

**¡Gracias Juan Spinelli
por la creación de parte de los
contenidos de este encuentro!**



ACÀMICA