

ACÀMICA

¡Bienvenidas/os a Data Science!

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



Agenda

¿Cómo anduvieron?

Repaso: SVM

Trade-off entre sesgo y varianza

Hands-On

Break

Cierre



Hoja de ruta

fase	ADQUISICIÓN Y EXPLORACIÓN		MODELADO				DEPLOY
	Exploración de datos	Feature Engineering	Machine Learning: Clasificación y Regresión	Optimización de parámetros	Procesam. del lenguaje natural	Sistema de recomendación	Publicación de modelos
tiempo	SEM 1	SEM 5	SEM 8	SEM 12	SEM 14	SEM 18	SEM 22
	SEM 2	SEM 6	SEM 9	SEM 13	SEM 15	SEM 19	SEM 23
	SEM 3	SEM 7	SEM 10		SEM 16	SEM 20	SEM 24
	SEM 4		SEM 11		SEM 17	SEM 21	



¿Cómo anduvieron?



Support Vector Machine (SVM): Resumen. Implementación.



SVM • Definición

Es un algoritmo de clasificación binaria que se propone encontrar el hiperplano que mejor separa los datos, tal que **maximice el margen**.

Hiperparámetro: C y Kernels.

`sklearn.svm.SVC`

« `class sklearn.svm. SVC (C=1.0, kernel='rbf', degree=3, gamma='auto_deprecated', coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=-1, decision_function_shape='ovr', random_state=None)`

[\[source\]](#)

SVM • Ventajas y desventajas

Ventajas

- Eficaz en espacios de alta dimensión (¡incluso cuando son más que el número de instancias!).
- Eficiente en memoria (sólo los vectores de soporte me definen el hiperplano frontera).
- Los kernels lo hacen súper versátil.

Desventajas

- Si usas kernels, es difícil no overfittear
- Funciona bien BIEN sólo para clasificación

SVM • Extensión

Extensión a múltiples clases: estrategia One vs. All

Para cada clase k , entrenar un SVM para discriminar k del resto .

Para una nueva instancia x ,
correr los K clasificadores
y retornar la clase para la cual x queda a mayor distancia del
hiperplano de decisión (i.e., la clase con mayor confianza).

SVM • Usos

- Para [datos linealmente separables](#), SVM funciona increíblemente bien.
- Para los [datos que son casi linealmente separables](#), SVM puede funcionar bien con el valor correcto de hiperplano.
- Para [datos que no son separables linealmente](#), podemos proyectar los datos al espacio donde es perfectamente o casi linealmente separable.

Actividad SVM: ¡A pensar!



¡A pensar!

¿A cuál de los siguientes ejemplos podemos aplicarle el algoritmo de SVM?

Caso 1: predecir el tipo de flor de acuerdo a las características dadas.

Caso 2: predecir el valor de una acción de la bolsa de valores, utilizando los valores históricos.

Caso 3: predecir si una persona puede vivir o morir en el hundimiento del Titanic tomando en cuenta la edad, el sexo y la ubicación de su cabina.



¡A pensar!

¿A cuál de los siguientes ejemplos podemos aplicarle el algoritmo de SVM?

Caso 1: predecir el tipo de flor de acuerdo a las características dadas. **CLASIFICACIÓN**



Caso 2: predecir el valor de una acción de la bolsa de valores, utilizando los valores históricos. **REGRESIÓN**



Caso 3: predecir si una persona puede vivir o morir en el hundimiento del Titanic tomando en cuenta la edad, el sexo y la ubicación de su cabina. **CLASIFICACIÓN**



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)

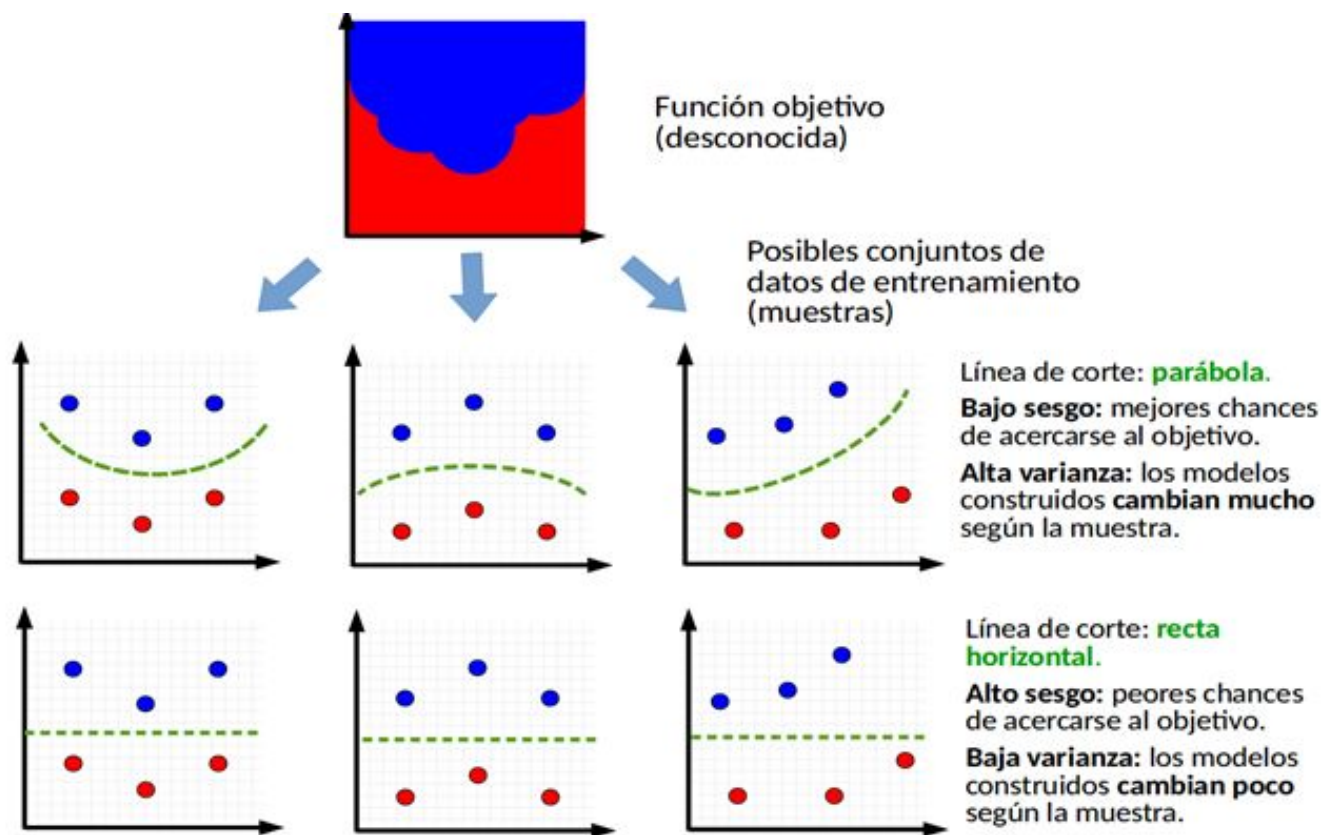
- No se puede reducir
- Se lo conoce como Ruido
- Proviene de factores como:
 1. Variables desconocidas que influyen en el mapeo de las variables de entrada a la variable de salida.
 2. Un conjunto de características incompletos
 3. Un problema mal enmarcado

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

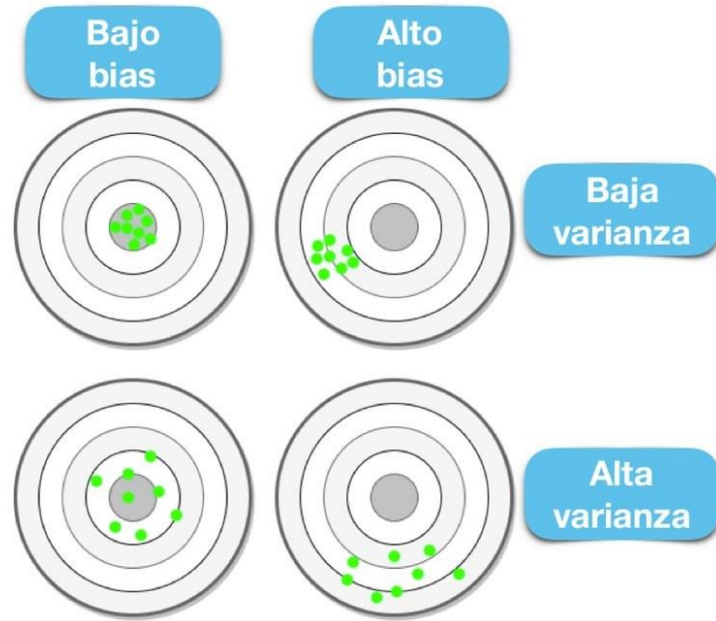
Regresión lineal, análisis discriminante lineal y regresión logística

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

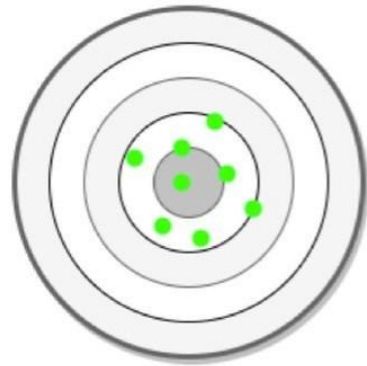


Baja varianza
Alto bias

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



Bajo bias
Alta varianza

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

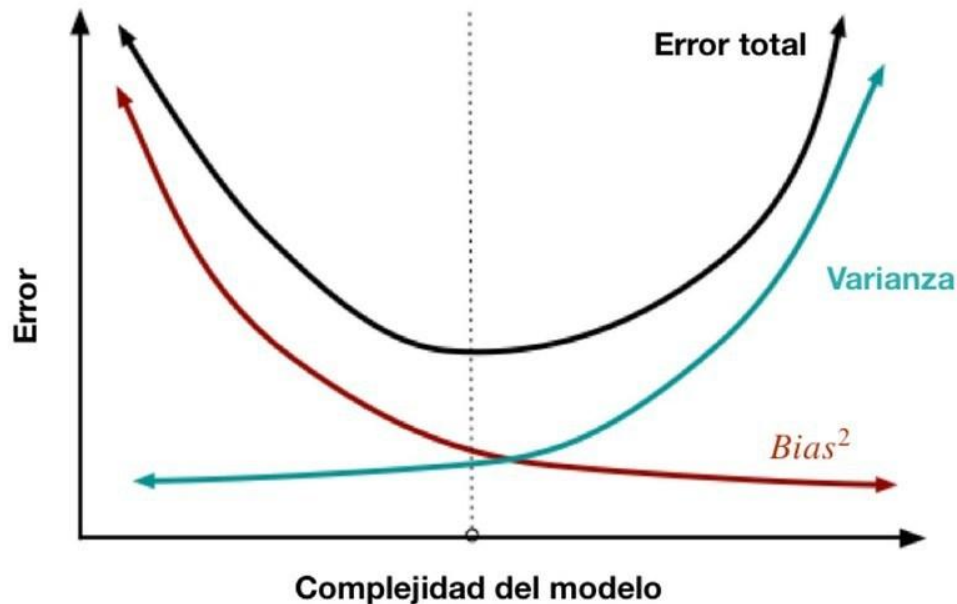
Entrenan modelos que son exactos en promedio pero inconsistentes. Estos incluyen algoritmos no lineales o no paramétricos, como árboles de decisión y k-vecinos más cercanos.

Entonces...

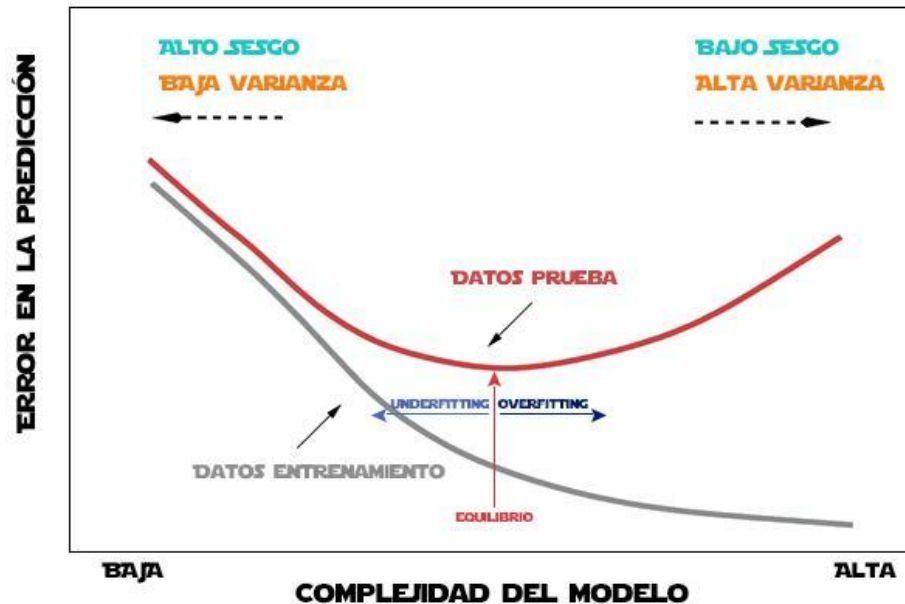
Cuanto más flexible sea nuestro modelo, la varianza aumentará, y el sesgo disminuirá.

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.



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.



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 sits on a matching white saucer. In the background, a white napkin and a silver spoon are partially visible, though out of focus. The overall lighting is soft and warm, creating a cozy atmosphere.

¡BREAK!

Ph. Credit: Drew Coffmann



Hands-on training



Hands-on training

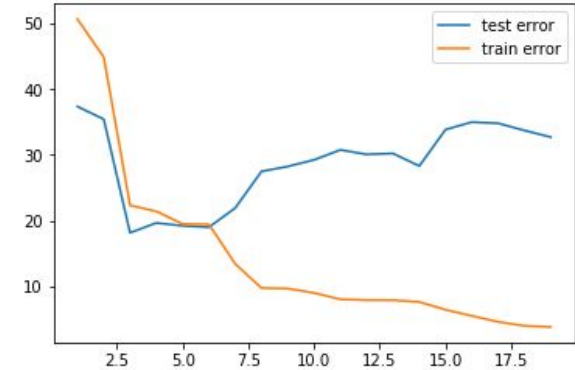
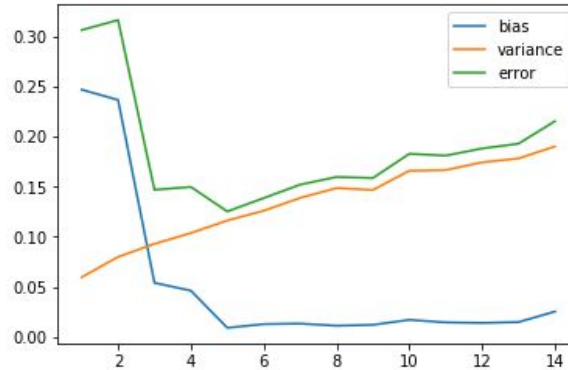
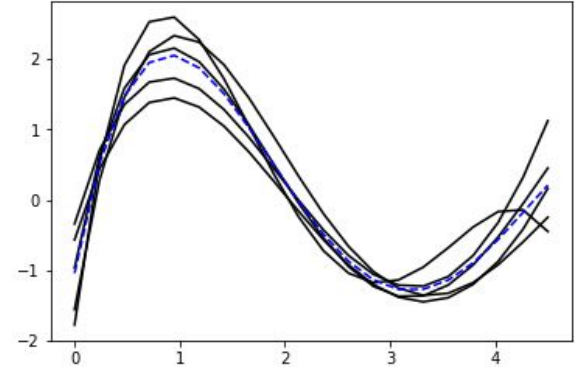
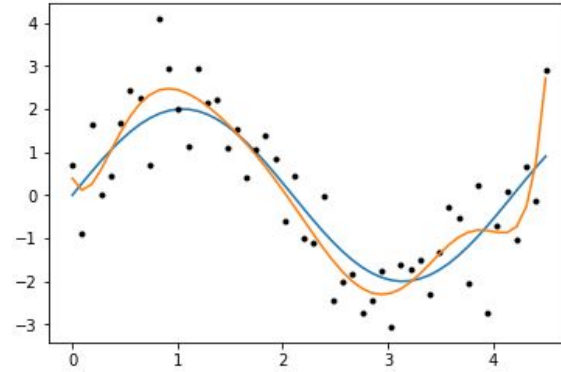


DS_Encuentro_26_Sesgo_Varianza.ipynb

bias_variance.ipynb

Notebook
adicional
de ejemplo

Sesgo y
varianza



Recursos



Si te quedaste con ganas de más...

De Tradeoff Sesgo-Varianza:

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

De SVM:

- [Support Vector Machines: A Visual Explanation with Sample Python Code](#)



Para la próxima

1. Ver los videos de la plataforma “Clasificación Avanzada: Ensamblés Bagging”.
2. Completar los notebooks de hoy y atrasados.

ACÀMICA