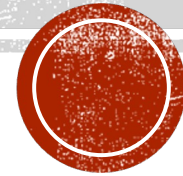
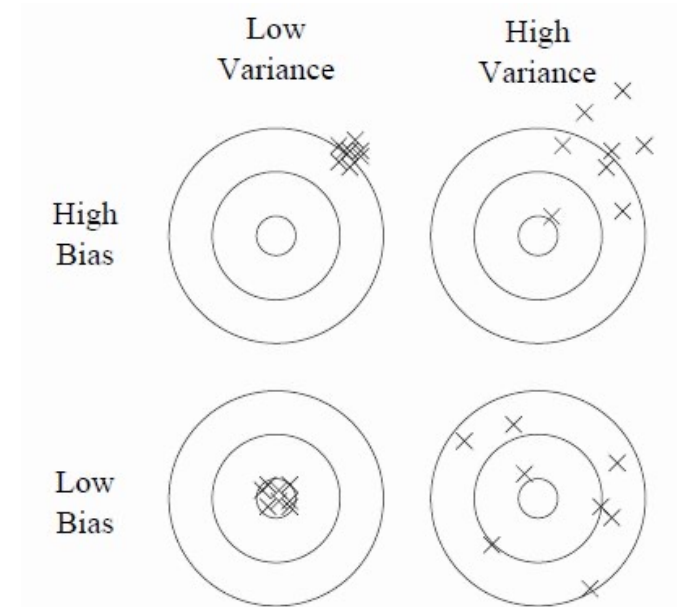


OVERFITTING Y PROTOCOLOS DE EVALUACIÓN



SESGO / VARIANZA

- **Sesgo** (bias): que tan lejos está el modelo de la verdad
- **Varianza**: Qué tanto varían los datos de la predicción para una misma instancia



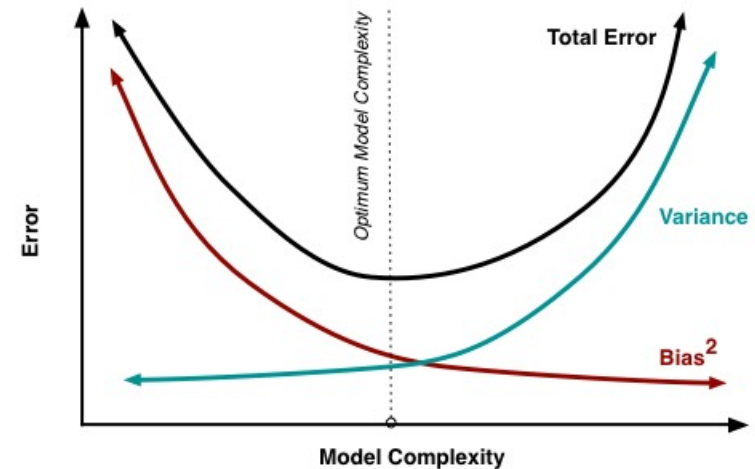
Domingo, 2012

$$Err(x) = \underbrace{\left(\overbrace{E[\hat{f}(x)]}^{\text{Promedio del modelo}} - \overbrace{f(x)}^{\text{Verdad}} \right)^2}_{\text{Sesgo}^2} + \underbrace{E \left[\overbrace{\hat{f}(x) - E[\hat{f}(x)]}^{\text{cada resultado del modelo}} \right]^2}_{\text{Varianza}} + \underbrace{\sigma_e^2}_{\text{Error irreducible}}$$



SESGO / VARIANZA

- Ambos son fuente de error
- Se debe determinar un **compromiso** entre ambos tipos de error
- Parámetros de cada modelo controlan la complejidad



<http://scott.fortmann-roe.com/docs/BiasVariance.html>



SOBRE APRENDIZAJE (OVERFITTING)

¿Cómo le enseño a mi hijo que es una pelota?

Set de entrenamiento



¿Qué patrones distinguen las pelotas de los demás juguetes?

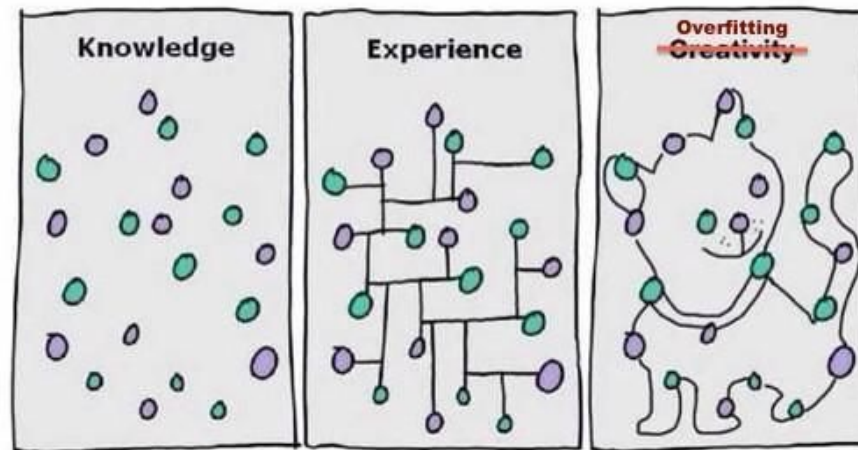
Hijo, ¿Es esta una pelota?



¿Cómo caracterizo una situación de modelos con overfitting?



SOBRE APRENDIZAJE (OVERFITTING)



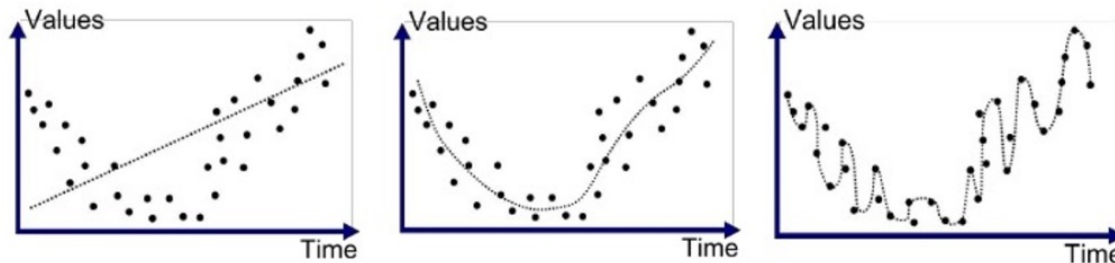
<http://blog.algotrading101.com/design-theories/what-is-curve-fitting-overfitting-in-trading/>

- **Sobre aprendizaje:** Los modelos aprenden a describir los errores aleatorios o el “ruido” del conjunto de entrenamiento.
- Ocurre cuando un modelo se vuelve excesivamente **complejo**

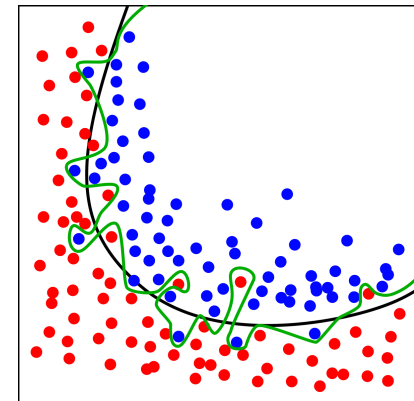


SOBRE APRENDIZAJE (OVERFITTING)

Regresión



Clasificación



¿Cómo es el sesgo y la varianza de estos modelos?

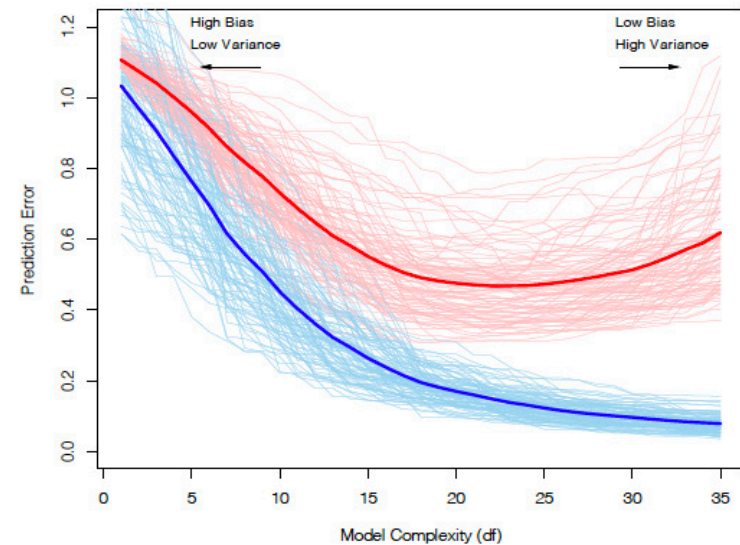
- La **complejidad** de un modelo debe ajustarse de tal manera que permita la **generalización**, al utilizarse con datos que no haya conocido durante el proceso de entrenamiento
- Principio de **parsimonia** (Occam's Razor): la mejor explicación es la más simple → preferir los modelos más sencillos con menos suposiciones

<https://en.wikipedia.org/wiki/Overfitting>



SOBRE APRENDIZAJE (OVERFITTING)

- Los modelos tienden a ajustarse al conjunto de datos de entrenamiento → el **error de entrenamiento** es un mal estimador
- Queremos encontrar la complejidad del modelo que nos permita minimizar el **error de prueba**



<https://onlinecourses.science.psu.edu/stat857/node/160>



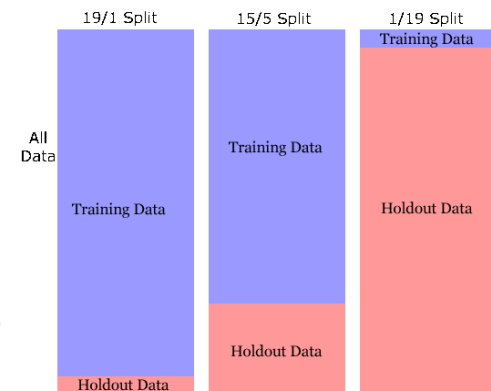
PROTOCOLOS DE EVALUACIÓN

- Aplican para aprendizaje supervisado en general (tanto para clasificación como para regresión).
- Permiten definir los mejores valores de los parámetros de los modelos en términos de **flexibilidad**.
- Evaluar cual sería la capacidad de **generalización** del modelo a datos nuevos
- Diferenciar entre el **error de entrenamiento** y el **error de test**. Evitar el sesgo causado por la **subestimación del error** al evaluar con el mismo set de entrenamiento.
- Permitir establecer un compromiso entre sesgo y varianza, luchando contra el **sobre aprendizaje**, en busca de un modelo con buenas **capacidades predictivas**

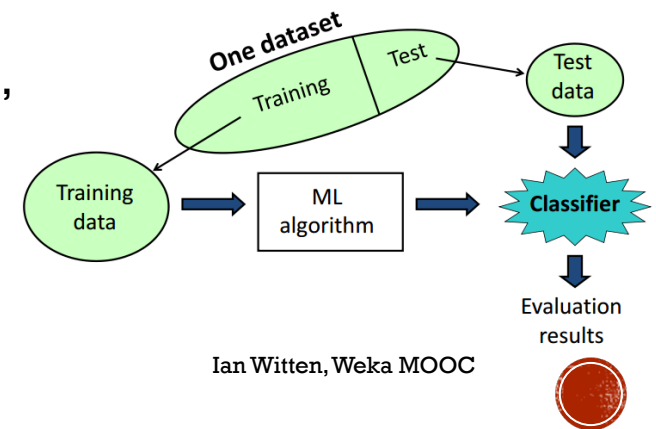


PROTOCOLOS DE EVALUACIÓN

- **Holdout:** particionar el conjunto de datos en 2:
 - **Conjunto de entrenamiento:** con el que se aprende el algoritmo de clasificación
 - **Conjunto de validación o test:** separa al comienzo del procedimiento y no se considera en el aprendizaje
 - **Aleatoriedad** del particionamiento
 - **Compromiso:** entre mas datos mejor el aprendizaje, entre mas datos mejor la evaluación
- **Repeated holdout:** repetir el procedimiento y agregar las métricas de evaluación



<https://webdocs.cs.ualberta.ca/~aixplore/learning/DecisionTrees/InterArticle/6-DecisionTree.html>

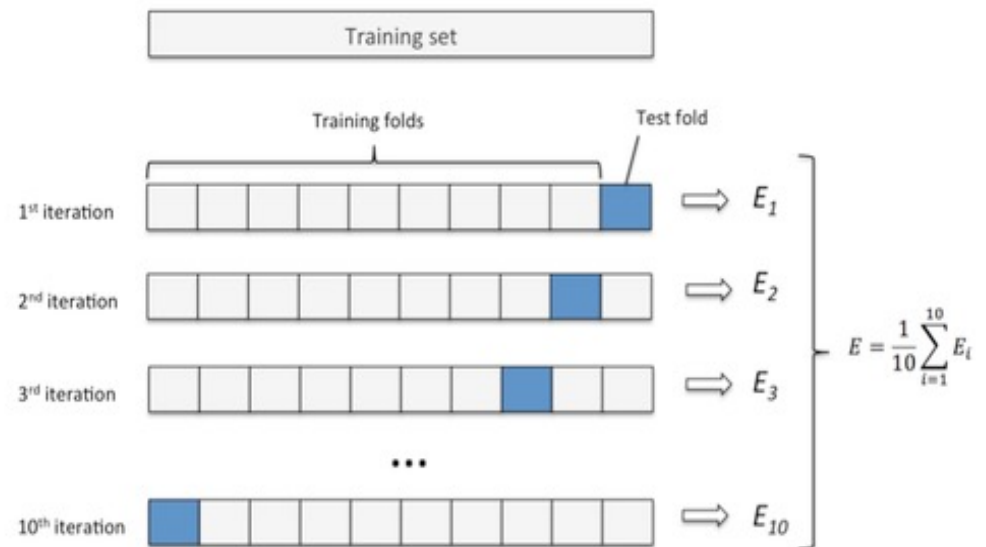


Ian Witten, Weka MOOC

PROTOCOLOS DE EVALUACIÓN

- **K-fold cross-validation:**

- Particionar el set de datos en K conjuntos disyuntos del mismo tamaño
- K-1 partes se usan para entrenamiento, 1 parte se usa para el test
- Se repite el proceso K veces
- Se agregan las métricas de evaluación



Sebastian Raschka, 2015

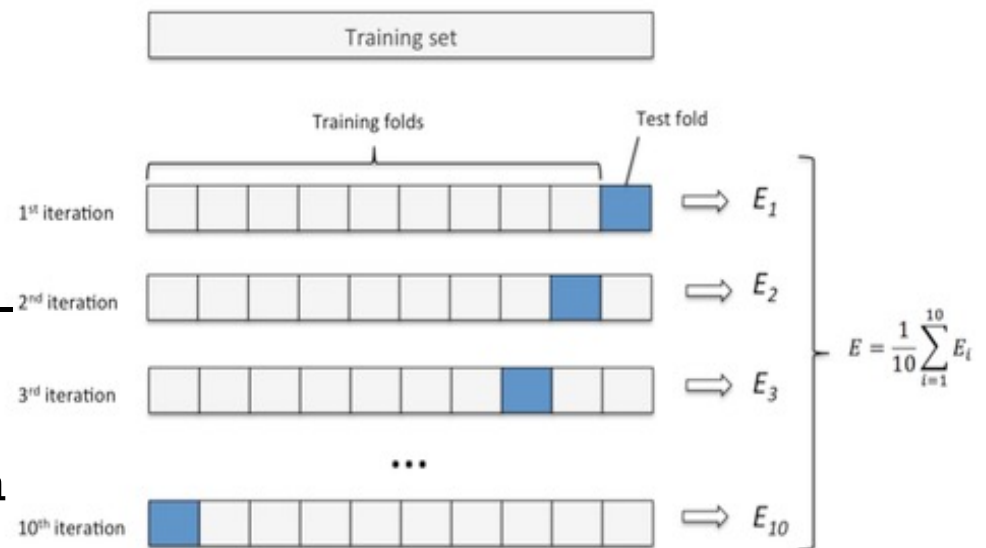


PROTOCOLOS DE EVALUACIÓN

- **K-fold cross-validation,**

Escogencia del K:

- Permite balancear entre sesgo y varianza
- **LOOCV** (Leave One Out Cross-Validation): partes de tamaño 1
- Por defecto se estima que los mejores resultados se obtienen con un valor de K entre 5 y 10



Sebastian Raschka, 2015



PROTOCOLOS DE EVALUACIÓN

- **Bootstrapping:**

- Consideración de varios conjuntos de entrenamiento/test utilizando muestreo con remplazo
- Por lo general muestreos del mismo tamaño del conjunto original
- Es un muy buen estimador de los parámetros, pero no de las métricas de calidad de los modelos dado un sesgo causado por el número de observaciones distintas promedio ($0.632 \cdot N$)

Original Dataset

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
-------	-------	-------	-------	-------	-------	-------	-------	-------	----------

Bootstrap 1

x_8	x_6	x_2	x_9	x_5	x_8	x_1	x_4	x_8	x_2
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Bootstrap 2

x_{10}	x_1	x_3	x_5	x_1	x_7	x_4	x_2	x_1	x_8
----------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Bootstrap 3

x_6	x_5	x_4	x_1	x_2	x_4	x_2	x_6	x_9	x_2
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Sebastian Raschka, 2015



PROTOCOLOS DE EVALUACIÓN

- **Set de validación vs set de test:**

- Separación de un set de datos de **test** para evaluación final del modelo escogido
- **Overfitting** si se calibran los modelos con el mismo set de test



- Calibración de pretratamientos (normalización, imputación)
- Calibración de los parámetros de los modelos (KFCV, holdout)
- Comparación de los resultados de diferentes modelos
- Evaluar la capacidad de generalización

