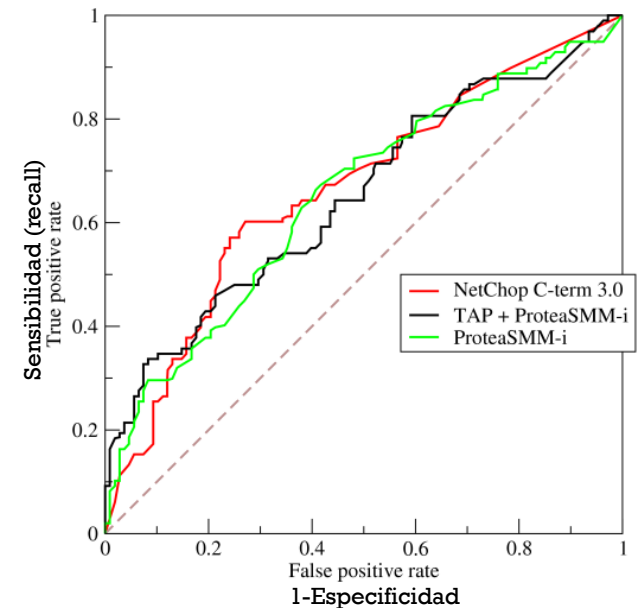


MÉTRICAS DE CLASIFICACIÓN

- **ROC AUC** (Area under the Curve)
 - **Receiver Operating Curve:** Curva de la tasa de los verdaderos positivos (recall o sensibilidad) contra los falsos positivos (1-especificidad)
 - Clasificadores binarios
 - Se hace variar el umbral de definición de predicción positiva y negativa
 - Permite establecer con los clientes posibles compromisos entre los dos tipos de errores



Wikipedia.org



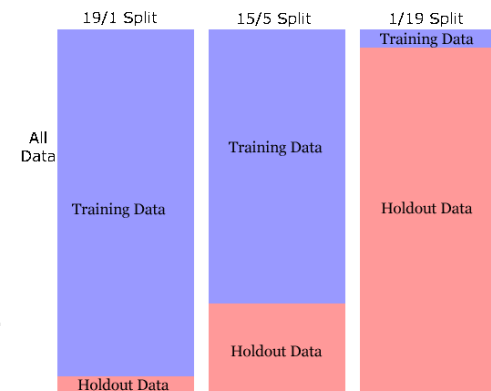
PROTOCOLOS DE EVALUACIÓN

- 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**

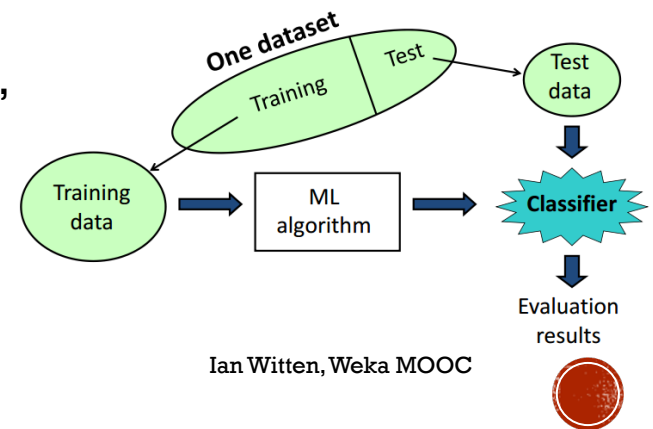


PROTOCOLOS DE EVALUACIÓN

- **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>

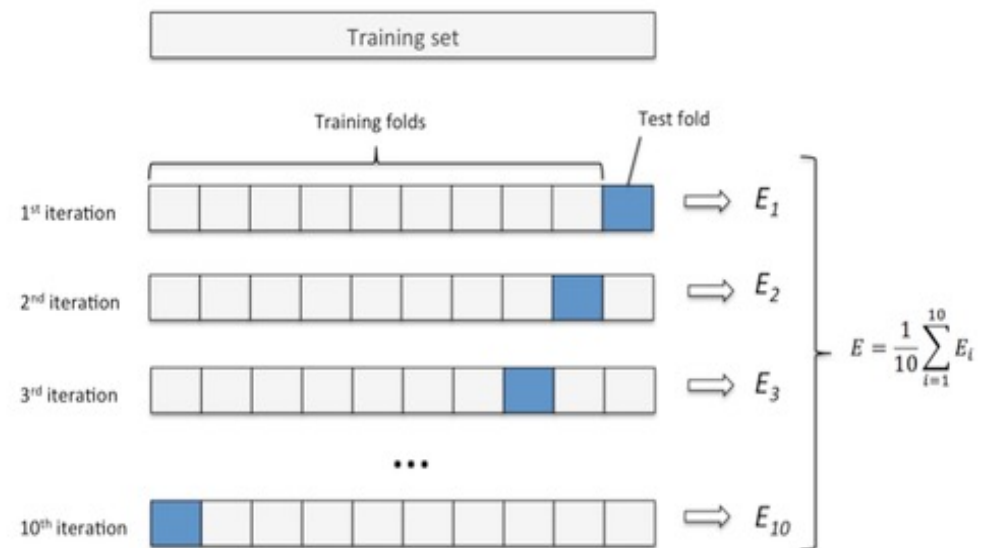


Ian Witten, Weka MOOC

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

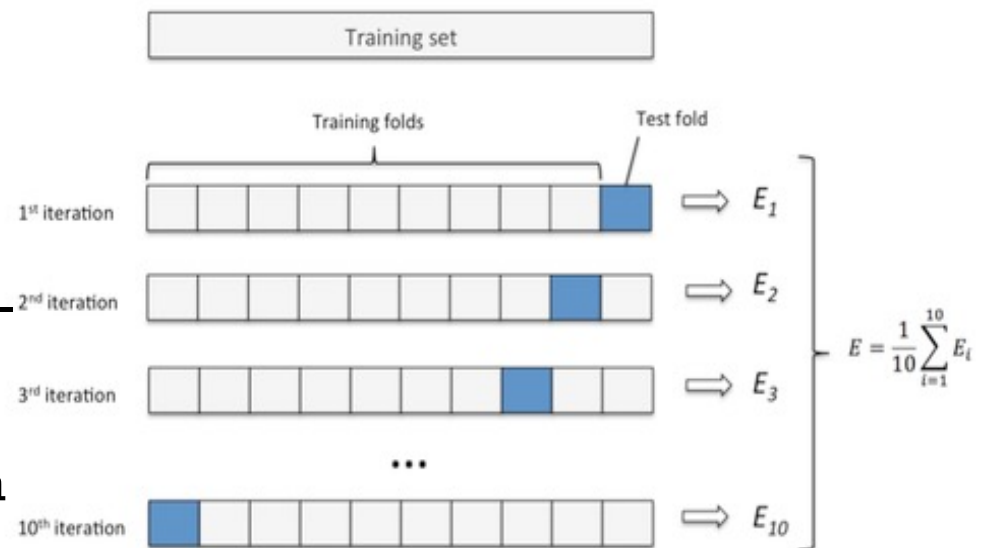


PROTOCOLOS DE EVALUACIÓN

- **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

