**¿En qué consiste un “protocolo de evaluación” y por qué es necesario?**

R/ En validar los modelos con pedazos del set de entrenamiento. Es necesario porque no se suele tener un set de validación en la vida real.

**Para simplicidad de la evaluación, ¿es válido evaluar la calidad de un modelo sobre el mismo set de entrenamiento? ¿Por qué?**

R/ No. Porque el modelo fue calculado con los mismos datos y con nuevos datos puede generar errores.

**Explique en que consiste el protocolo de hold-out y si vale la pena repetirlo para evaluar un modelo de aprendizaje supervisado.**

R/ Consiste en dividir aleatoriamente el total de observaciones en 2 grupos. Un set de entrenamiento y un set de validación (hold-out).

Se obtiene el modelo con el set de entrenamiento y luego se usa ese modelo para predecir respuestas para las observaciones del set de validación.

**Explique en que consiste el protocolo de LOOCV y que implicaciones tiene frente al holdout en un contexto de Big Data.**

R/ Solo se deja una observación para el set de validación. El proceso se repite n veces siendo n la cantidad de observaciones. Esto da n MSE los cuales se promedian y dan un estimado del test MSE. Las implicaciones son: Menos bias, no sobreestima el test MSE, mucho menos variable, pesado de implementar puesto que requiere obtener n modelos.

**Explique en que consiste el protocolo de K-Fold cross validation y por qué es preferible a uno de hold-out.**

R/ Se divide aleatoriamente el total de observaciones en K grupos. Cada grupo K se utiliza como set de validación y se generan k modelos. El método de arriba es este con K=n.

Es preferible por lo mismo que LOOCV.

**Expliquen la diferencia entre sesgo (bias) y varianza con respecto a su contribución en el error de predicción.**

R/ Sesgo: Error inherente de los datos.

Varianza: Error de la predicción.

**Explique en que consiste el protocolo de Bootstrap**

R/ Algo así como… Obtener varios subsets aleatorios del total de datos. Calcular a para cada subset. Y luego sacar el promedio de todas las a. a es el valor que minimiza la Varianza.