

Pronóstico de la demanda en empresas retail

Técnica basada en Business Intelligence y Machine Learning

Raúl Benítez - Alberto Garcete

Tutores: PhD. Diego P. Pinto Roa - Ing. Aditardo Vázquez

Universidad Nacional de Asunción - Facultad Politécnica

Agosto 2018



Complejidad del modelo

La capacidad de generalización está determinada principalmente por dos factores:

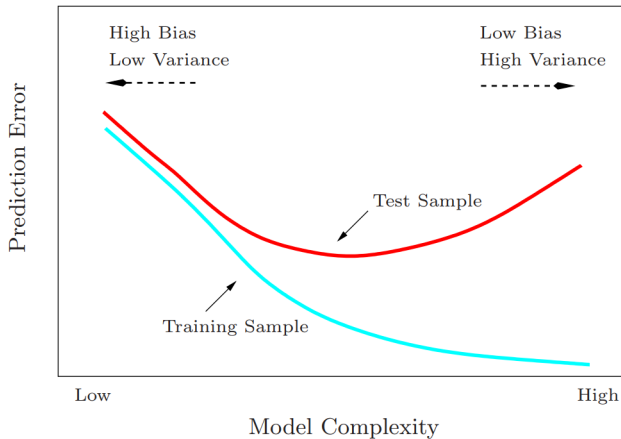
- La complejidad del modelo (model complexity).
- El tamaño del conjunto de entrenamiento (learning curves).

Muestra la relación existente entre la complejidad del modelo (eje x) y el error de predicción (eje y). Indica cómo va variando el error de predicción en los conjuntos de entrenamiento y de prueba a medida que aumenta la complejidad del modelo. Los modelos muy simples dan una situación no deseada de sesgo (subajuste) y los modelos muy complejos igualmente una situación no deseada de varianza (sobreajuste). Para obtener una buena capacidad de generalización se debe encontrar un compromiso entre sesgo y varianza.

Este compromiso entre sesgo y varianza se puede encontrar probando distintas configuraciones de los parámetros de ajuste (tuning parameters) de los algoritmos de aprendizaje y luego eligiendo el que arroje mejores resultados. Normalmente se recomienda probar las distintas configuraciones de los parámetros sobre un conjunto de datos llamado de validación, que no se debe confundir con los conjuntos de entrenamiento y de prueba.



Complejidad del modelo



Complejidad del modelo

Algunos ejemplos de familias de algoritmos y sus parámetros de ajuste son:

- Algoritmos de regresión lineal: la cantidad de atributos del modelo (que determinará la cantidad de variables independientes).
- Modelos bayesianos: en las redes bayesianas por ejemplo se debe elegir el algoritmo estimador (estimación de las tablas de probabilidad condicional de las redes bayesianas) y el método para encontrar estructuras de red.
- Árboles de decisión: el factor de confianza de la poda, mínimo número de instancias, número de pasadas para reducir el error de poda [1].
- Redes neuronales artificiales: la topología de la red donde se debe elegir la cantidad de neuronas para la capa de entrada, salida y la oculta. También se deben probar distintas tasas de aprendizaje (learning rate) y funciones de activación.
- Support vector machines: los parámetros gamma (γ), costo (C), y epsilon (ϵ). γ es una constante que reduce el espacio modelo y controla la complejidad de la solución, C es una constante positiva que es un parámetro de control de capacidad, mientras que ϵ es la función de pérdida que describe el vector de regresión sin todos los datos de entrada [2].



Curvas de aprendizaje

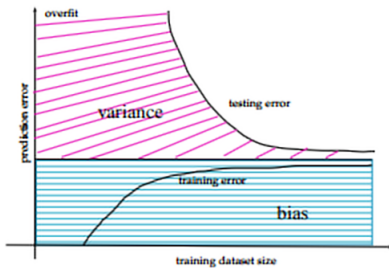
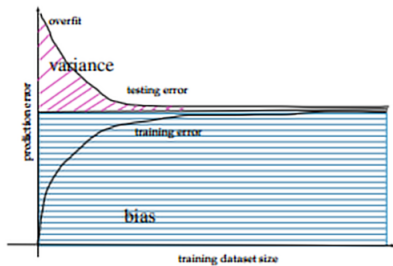
Muestra la relación existente entre el tamaño del conjunto de entrenamiento (eje x) y el error de predicción (eje y). Indica cómo va variando el error de predicción en los conjuntos de entrenamiento y de prueba a medida que aumenta el tamaño del conjunto de entrenamiento. Los modelos muy simples dan problema de alto sesgo y los modelos muy complejos dan problema de alta varianza.

Cuando el modelo está sesgado por mas que se aumenten los datos del conjunto de entrenamiento el nivel de error del conjunto de entrenamiento y el de prueba se mantienen alto, por lo que para modelos sesgados no es una solución aumentar significativamente la cantidad de los datos del conjunto de entrenamiento. La solución se debería buscar aumentando la complejidad del modelo.

Cuando el modelo tiene alta varianza el error en el conjunto de entrenamiento se mantiene bajo aún aumentando su cantidad de datos. El conjunto de prueba en cambio inicia con alto porcentaje de error con pocos datos, pero a medida que aumentan los datos en el conjunto de entrenamiento el error del conjunto de prueba disminuye significativamente. Por tanto, ante alta varianza es una posible solución aumentar el tamaño de los datos. También se puede probar disminuyendo la complejidad del modelo.



Curvas de aprendizaje





R. G. Mantovani, T. Horváth, R. Cerri, J. Vanschoren, and A. C. P. L. F. d. Carvalho.

Hyper-parameter tuning of a decision tree induction algorithm.

In *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 37–42, Oct 2016.



Soukayna Mouatadid and Jan Adamowski.

Using extreme learning machines for short-term urban water demand forecasting.

Urban Water Journal, 14(6):630–638, 2016.

