

ISLR. Capítulo 2. Ejercicios

Javier Santibáñez

23 de agosto de 2015

Conceptuales

1. Para cada una de las partes a) a d), indicar si podemos esperar o no que el desempeño de método de aprendizaje estadístico flexible sea mejor o peor que un método inflexible. Justifique su respuesta.
 - a) El tamaño de muestra n es extremadamente grande y el número de predictores p es pequeño.
 - b) El número de predictores es extremadamente grande y el número de observaciones n es pequeño.
 - c) La relación entre los predictores y la respuesta es altamente no lineal.
 - d) La varianza de los errores $Var(\epsilon) = \sigma^2$ es extremadamente grande.

Respuesta

- a) Pendiente.
 - b) Pendiente.
 - c) Será mejor el desempeño de un método flexible. Como la relación es marcadamente no lineal, usar un método inflexible nos daría un sesgo cuadrado elevado al restringir la forma del modelo.
 - d) Será mejor el desempeño de un método inflexible. Como la variación intrínseca a los datos es elevada, usar un método flexible podría provocar un sobreajuste, lo cual estaría asociado con una elevada varianza del modelo (varianza debido a que el modelo se entrena con una muestra).
2. Explicar en cada escenario si se trata de un problema de regresión o de clasificación, además indicar si se está más interesado en inferencias o predicciones. Finalmente, indique los valores de n y p .
 - a) Colectamos datos de las 500 firmas *top* de EUA. Para cada firma registramos ganancias, número de empleados, industria y el salario del CEO. Estamos interesados en entender qué factores afectan el salario del CEO.
 - b) Consideramos el lanzamiento de un nuevo producto y deseamos saber si será un *éxito* o un *fracaso*. Colectamos datos de 20 productos similares que fueron lanzados previamente. De cada producto registramos si resultó en éxito o fracaso, precio por cargo del producto, presupuesto de marketing, precio de la competencia y otras 10 variables.
 - c) Estamos interesados en predecir el % de cambio en el dólar de EUA en relación a los cambios semanales en las bolsas mundiales. Entonces colectamos semanalmente datos de todo el año. En cada semana registramos el % de cambio en el dólar, el % de cambio en el mercado de EUA, el % de cambio en el mercado británico y el porcentaje de cambio en el mercado alemán.

Respuesta

- a) Se trata de un problema de regresión porque la respuesta (salario del CEO) es cuantitativa. Es un problema de inferencia porque queremos entender las relaciones de los predictores con la respuesta. $n = 500$ y $p = 3$.
- b) Se trata de un problema de clasificación porque la respuesta es binaria (éxito o fracaso). Se puede considerar como un problema de inferencia porque al entender las relaciones de los predictores y el éxito del producto, se puede formular una estrategia para asegurar un lanzamiento exitoso. $n = 20$ y $p = 13$.
- c) Se trata de un problema de regresión porque la respuesta (% de cambio del dólar) es cuantitativa. Es un problema de predicción ya que la teoría económica establece que relaciones hay entre los predictores y la respuesta; además, al no poder controlar los cambios en los predictores la utilidad principal del modelo será la predicción.