

Taller de Consultoria - TP5

Gonzalo Barrera Borla

13/11/2019

Setup

```
library(tidyverse) # manipulación de datos en general, graficos
library(broom) # limpieza y estructuración de resultados de regresiones
library(caret) # Validación cruzada para múltiples modelos
```

Ejercicio 2

Los siguientes datos corresponden a un trabajo para determinar la composición de un conjunto de vasijas de vidrio de un yacimiento arqueológico. Como el análisis espectrométrico es más barato que el análisis químico, se procuró calibrar el primero para que reemplace al segundo. Con este objetivo se tomó una muestra de 180 vasijas, a las que se realizó una espectrometría de rayos X sobre 1920 frecuencias, y también un análisis de laboratorio para determinar el contenido de 13 compuestos químicos, a saber:

Na₂O MgO Al₂O₃ SiO₂ P₂O₅ SO₃ Cl K₂O CaO MnO Fe₂O₃ BaO PbO

Cada fila del archivo Vessel_X es el espectro de una vasija, limitado a las frecuencias 100 a 400, pues las demás tienen valores casi nulos. Cada fila del archivo Vessel_Y tiene los contenidos de los 13 compuestos en esa vasija. Se trata de predecir el compuesto 1 (óxido de sodio) usando sólo las columnas 10, 20, ... etc. de X.

1. Para familiarizarse con los datos, grafique en función de la frecuencia las medias y varianzas de X, y también algunos espectros.
2. Luego tome una muestra al azar de 120 vasijas. Con ella calcule los estimadores de mínimos cuadrados, Forward y Backward; y cualquier otro que se le ocurra. Para cada estimador estime el error cuadrático medio de predicción (ECM).
3. Luego aplíquelos a las otras 60 vasijas para estimar el error de predicción (este será insesgado). Compare los estimadores, y también compare las estimaciones del ECM con el inicial.