

Laboratorio 4

Estadística Computacional

Universidad Técnica Federico Santa María
Departamento de Informática

Héctor Allende Olivares
<hallende@inf.ut fsm.cl>

Erick López Ovando
<elopez.ovando@gmail.com>

Camilo Valenzuela Carrasco
<camilo.valenzuela@alumnos.usm.cl>

16 de Diciembre del 2015

1. Objetivo

En este laboratorio, el alumno aprenderá a realizar modelos de **Regresión Lineal Múltiple**¹. Tendrá que poder reconocer cuándo un modelo es mejor a otro, y además se utilizará métodos de selección de variable, para simplificar el modelo, tratando de mantener su calidad.

2. Motivación

Ya estamos a Diciembre, se acerca la Navidad y las Fiestas de Fin de Año. Alex Simmons, el compañero de trabajo de los ayudantes, reconocido por su amor al vino en caja, ha decidido mejorar la calidad de su brebaje favorito. Por lo que le ha pedido a los alumnos de Estadística Computacional encontrar las características más relevantes que inciden en la calidad del Vino y un modelo de regresión lineal, para poder saber con anticipación si el vino que va a comprar va a ser de calidad o no.

3. Datos a utilizar

Los datos que serán utilizados en este laboratorio son el Wine Quality Data Set², del UCI Repository.

Este dataset tiene 4898 mediciones, que están divididos en dos sets, uno del vino tinto y otro del vino blanco (Tiene que utilizar ambos para generar los modelos). Cada medición tiene 11 variables y lo que se busca es poder predecir la calidad.

¹https://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal

²<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

4. Preguntas

Para todos los ejercicios que lo necesiten utilice como criterio de selección $\alpha = 0,05$

- 1.- Utilizando todas las variables, genere un modelo de Regresión Lineal. ¿Que tan bueno es el modelo? Comente sus resultados.
- 2.- Utilizando Eliminación Hacia Atras "Backward", genere el modelo de Regresion Lineal. Muestre paso a paso que variable se eligen, porqué y cómo va mejorando o empeorando el Modelo.
- 3.- Utilice Selección Hacia Adelante "Forward", genere otro modelo de Regresión Lineal, mostrando todos los pasos realizados y como va cambiando el modelo.
- 4.- Realice el mismo ejercicio que en los puntos anteriores, pero ahora utilizando la combinación de Backward y Forward (Stepwise).
- 5.- Realice una comparación entre los modelos finales de cada método y el modelo inicial y concluya.
- 6.- ¿Que tan bueno son los modelos de Regresión Lineal? ¿Cuándo conviene utilizar Selección de Variables?

5. Sobre la Entrega

- 1.- El informe debe ser realizado en \LaTeX .
- 2.- Estructura del informe: El informe tiene que contener lo siguiente:
 - 2.1.- Introducción.
 - 2.2.- Desarrollo.
 - 2.3.- Conclusiones generales y conclusiones relacionadas al objetivo del laboratorio.
 - 2.4.- Referencias, si es que utilizaron. (Tienen que estar citadas en el informe).
- 3.- Los calculos y gráficos deben realizarse utilizando R, este código no tiene que estar incluido dentro del informe.
- 4.- El informe y el código utilizado tiene que ser subido a moodle en formato nombreCompleto1-rol1-nombreCompleto2-rol2.tar.gz.
- 5.- Fecha de entrega: **Miércoles 23 de Diciembre**. Se descontarán 20 puntos por cada día de atraso.