

Tarea 1. Introducción a Analítica. Escuela de Estadística. Facultad de Ciencias. Universidad Nacional de Colombia, Sede Medellín

Modo y fecha de entrega: La entrega de las tareas es por medio electrónico, vía email y se deben enviar a jcsalaza@unal.edu.co el sábado 22 de agosto de 2020 hasta las 6:00pm; se usará el reloj de mi PC para llevar el registro de entrega. Tareas que se entreguen entre las 6:01 y 6:30pm se calificarán sobre 4:00, y tareas que se entreguen después de las 6:30pm se calificarán como reprobadas. Se pueden conformar grupos de máximo 4 personas.

1. (10 pts. Teórico) Considere un estimador \hat{f} y un input X con los que se obtiene la predicción $\hat{Y} = \hat{f}(X)$. Asuma que \hat{f} y X son fijos. Demuestre que:

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(X) + \varepsilon - \hat{f}(X)]^2 \\ &= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{Reducible}} + \underbrace{\text{var}(\varepsilon)}_{\text{Irreducible}} \end{aligned}$$

2. (25 pts. Teórico) Es posible demostrar que la tasa de error

$$\text{Average}(I(y_0 \neq \hat{y}_0))$$

se minimiza en promedio por medio de un clasificador muy simple que clasifica cada observación a la clase más probable o factible, dados los valores de sus predictores. En otras palabras, se asigna simplemente una observación del conjunto de prueba con predictor x_0 a la clase j para la cual

$$Pr(Y = j | X = x_0)$$

sea la más grande. Esta es una probabilidad condicional. Este clasificador tan simple se conoce como CLASIFICADOR DE BAYES. Demuestre que el clasificador de Bayes produce la menor tasa de error de prueba (Test Error Rate) posible y se conoce como [Tasa de Error de Bayes](#).

3. (20 pts. Teórico y práctico) Ejercicio 7, texto guía página 53. Hágalo también en R.
4. (45 pts. Práctico) Ejercicio 8, texto guía página 54. Los datos se cargan con la librería de R llamada [ISLR](#) y la instrucción `college=ISLR::College`