Tarea 6 Reconocimiento de Patrones

Entregar a más tardar el domingo 19 de mayo 10PM.
Se debe integrar todo en un solo pdf (sin comprimir).

Echa un ojo a los dos colabs con el código para ajustar árboles usando Python y R. Sugiero echar un ojo primero al de R.

1. (no entregar nada) Lee sección 8.1 del libro *Introduction to Statistical Learning* (with R o with Python): ver `https://www.statlearning.com/`

   Otra referencia general (mucho más allá del curso) es:
   `https://hbiostat.org/rmsc/lrm`

2. Sea $Y \sim Bern(0.5)$, $X|Y=0 \sim \mathcal{N}((0,0), \mathbb{A})$ y $X|Y=1 \sim \mathcal{N}((1,2), \mathbb{A})$ con

$$\mathbb{A} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

   a) Calcula el clasificador Bayesiano óptimo suponiendo una función de costo simétrica.

   b) Genera muestras (X, Y) de tamaño $n = 50, 100, 500$. Visualiza las muestras.

   c) Busca y discute árboles de decisión adecuados para estos datos y compara los mejores modelos obtenidos con validación cruzada con el clasificador Bayesian óptimo, tanto de manera visual como de manera cuantitativa.
   Muestra que podar el árbol puede aumentar el poder predictivo.

   d) Ajusta un modelo de regresión logistica con (solamente) efectos principales (i.e. $y \sim x_1 + x_2$) para estos datos; discute su ajuste. Compara su poder predictivo con un modelo de regresión logistica con términos cruzados (i.e. $y \sim x_1 + x_2 + x_1 : x_2$)

   e) Ajusta clasificadores SVM; con kernel lineal y kernel de base radial. Estima su poder predictivo y compáralo con lo obtenido en los incisos anteriores.

3. Busca algunos árboles de decisión, SVM y modelos de regresión logistica para los datos `Wisconsin breast cancer` database sobre tumores malignos ($Y = 0$) y benignos ($Y = 1$). Ver
`https://search.r-project.org/CRAN/refmans/mlbench/html/BreastCancer.html`
(La última columna indica el valor de Y) o
`http://archive.ics.uci.edu/dataset/14/breast+cancer`
Puedes quitar las observaciones con valores incompletos (NA).

   Evalua la calidad de ajuste.

4. (después de la clase de miércoles) Mira el video `https://youtu.be/bpZC-2zkOzk`. Motiva como el término con la función hinge se relaciona con el clasificador Bayesiano Optimo.

   Usando un camino similar (pero incluso más sencillo), verifica que se obtiene el mismo resultado para $E\exp(-Yg(X))$, o sea, usando en la función de costo el término $\sum_i \exp(-y_i g(x_i))$.

   Nota: no se debe demostrarlo pero lo anterior también aplica para $E(1 - Yg(X))^2$.

5. (para pensar) Hasta ahora estamos para todos los modelos de predicción suponiendo que tenemos una muestra de (X, Y). Desafortunadamente, eso no corresponde a lo que hacemos en la práctica.

   Piensa en un problema médico donde la v.a. binaria Y indica si una persona sobrevivió una enfermedad o no. Una manera para obtener datos es tomar cierta cantidad de personas muertas y medir su X, y como control tomar otra cantidad de personas que sobrevivieron. Eso se llama un muestreo *retrospectivo*. Contrástalo con un muestreo *prospectivo* donde seguimos personas en el tiempo hasta observar Y, este muestreo es en general lo ideal (pero mucho más costoso).

   En aprendizaje máquina muchas veces se trabaja con muestras *retrospectivos*. Por ejemplo para construir un clasificador de correos en spam o no spam, se toma cierta cantidad correos spam y cierta cantidad nospam. Convéncete que eso no es lo mismo que tomar una muestra de (X, Y).

   En RL estimamos $P(Y = y|X = x)$ a partir de una muestra $\{(x_i, y_i)\}$. Un estudio retrospectivo se puede modelar con una v.a. Z que indica si un dato entra a la muestra o no. Entonces lo que se está estimando es

$P(Y = y|X = x, Z = 1)$. Usando la regla de Bayes (un buen ejercicio) se obtiene:

$$P(Y = y|X = x, Z = 1) = \frac{P(Z = 1|X = x, Y = y)P(Y = y|X = x)}{\sum_j P(Z = 1|X = x, Y = j)P(Y = j|X = x)}$$

Si se hace el supuesto que $P(Z = 1|X = x, Y = y)$ no depende de x (es decir, dentro de la subpoblación con $Y = y$ se toma la submuestra independiente de x), entonces si definimos $\rho_y := P(Z = 1|Y = y)$ y usamos un modelo de RL para $P(Y = y|X = x)$ se obtiene (un buen ejercicio):

$$P(Y = y|X = x, Z = 1) = \frac{\rho_1 \exp(\alpha + \beta x)}{\rho_o + \rho_1 \exp(\alpha + \beta x)}$$

Entonces $log(P(Y = 1|X = x, Z = 1)/P(Y = 0|X = x, Z = 1)$ es de la forma $\alpha^* + \beta x$ con $\alpha^* = \alpha + log(\rho_1/\rho_0)$. Vemos que obtenemos con la muestra retrospectiva la misma β que con el modelo de RL para $P(Y = y|X = x)$ y eso es lo que es nos interesa. Es una gran ventaja que con un modelo RL bajo el supuesto que $P(Z = 1|X = x, Y = y)$ no depende de x, se puede obtener información sobre el efecto de los predictores aún cuando la muestra es retrospectiva.

(Lo anterior está basado en sección 5.1.4 del libro de Agresti *Categorical Data Analysis*).