Bayesian Statistics and Machine Learning Workshop 2023

Introuction to Classification Methods

Martín Onetto

1. Questions

1. Cómo es el proceso de clasificación?
2. Decribir el proceso que se definen las fronteras de decisión para la regresión lineal, logistica, LDA y QDA
3. ¿Cómo se podría *bayesianizar* los métodos vistos?
4. ¿Qué cambiaría en los resultados obtenidos?
5. ¿Cómo afectaría el número de datos de cada clase a las fronteras de las etiquetas?

1.1. Problems

Gaussians

Simular datos de 3 Gaussianas multivariabdas con $\Sigma = diag(1,1)$ y $\mu_1 = (0,0)$,$\mu_2 = (1,1)$ y $\mu_3 = (-1,-1)$. Cada guassiana ahora representa una clase $Y = k$ con $k \in \{1,2,3\}$

1. Armar las fronteras de decisión con cada uno de los métodos vistos.
2. Para cada método evalular el Error rate de la clasiicación.
3. Encontrar la dirección de máxima varianza según el método de Fisher y hacer un plot de las gaussianas sobre esa dirección.

Vowel recognition

Cargar la base de datos vowel.csv que contiene una variable target $y = k \in \{1, \ldots, 11\}$. Cada y_k corresponde a una vocal estas son:

vowel	word	vowel	word
i	heed	O	hod
I	hid	C:	hoard
E	head	U	hood
A	had	u:	who'd
a:	hard	3:	heard
Y	hud		

Para determinar estas vocales tenemos datos X con dimensión $p = 10$ que corresponden a procesamientos auditivos de diferentes frecuencias de mediciones de voz. Para este conjunto de datos:

1. Armar las fronteras de decisión con regresión logistica, LDA y QDA.
2. Para cada método evalular el Error rate de la clasiicación.
3. Encontrar las dos direcciones de máxima varianza según el método de Fisher y hacer un scatter plot de los datos con sus respectivas clases en esas direcciones.
4. Para el caso de la regresión logistica, calcular la posterior de los parámetros y su varianza. Con al información de la posterior queremos ver cómo se ve afectada la linea de decisión respeto a usar el máximun likelihoo. Para esto elegir dos clases g_i y g_j y calcular su frontera marginalizando sobre la posterior de β es decir:

$$\log \frac{P(G = i|X)}{P(G = j|X)} = \frac{\int P(G = i|X, \beta)P(\beta|X)d\beta}{\int P(G = j|X, \beta)P(\beta|X)d\beta} \tag{1}$$

   tomar como posterior de β su aproxmiación gaussiana por el método de Laplace.
5. Trazar las curvas fronteras entre las dos clases elegidas que resultan de samplear 1000 β^s de la posterior inferida.
6. Discutir la diferencia entre considerar la incerteza en los parámetros y no hacerlo.
7. Repetir el procedimiento anterior para QDA pero sólo hacer inferencia bayesiana sobre los μs, es decir seguir estimando cada matriz de covarianza como $\Sigma_k = \frac{1}{N_k - p - 1} \sum_{i=1}^{N_k} (y_i^k - \bar{y}^k)(y_i^k - \bar{y}^k)^T$.