

PRÁCTICA CALIFICADA 1

1. Muestre que una variable aleatoria Gaussiana tiene una distribución de probabilidad perteneciente a la familia de distribuciones exponenciales. (5 puntos)
2. Considere el método de clasificación KNN con $K = 3$, la norma $\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} \{|x_i|\}$, y los siguientes datos

Obs.	Data	X_1	X_2	X_3	Y
1	train	1/2	1	3/4	Verde
2	train	1	1	1/4	Verde
3	train	3/4	1/2	3/4	Verde
4	train	1/4	1/4	1/2	Azúl
5	train	1	3/4	1/4	Rojo
6	train	1/2	1/2	3/4	Azúl
7	test	0	0	0	Rojo
8	test	1/2	1/2	1/2	Azúl
9	test	1	1	1	Verde

- (a) Clasifique las observaciones en data de prueba. (2 puntos)
 - (b) Calcule el ratio de error de su clasificación en (a). (1 punto)
3. Considere el caso de regresión lineal múltiple de $\mathbf{y} \in \mathbb{R}^n$ con respecto a $\mathbf{X} \in \mathbb{R}^{n \times p}$
 - (a) Muestre que la recta de regresión estimada siempre pasa por el punto $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. (2.5 puntos)
 - (b) Muestre que el R^2 es igual al cuadrado de la correlación entre \mathbf{y} y $\hat{\mathbf{y}}$. (2.5 puntos)
 - (c) Muestre que $\mathbf{H}_{ii} \in [0, 1]$, donde $\hat{\mathbf{f}} = \mathbf{H}\mathbf{y}$. (2.5 puntos)
 - (d) Muestre que $\sum_{i=1}^n \mathbf{H}_{ii} = p$. (2.5 puntos)
 4. Considere el caso de regresión lineal de $\mathbf{y} \in \mathbb{R}^n$ con respecto a $\mathbf{x} \in \mathbb{R}^n$. Suponga que se recogen datos y se separan en data de entrenamiento (70% de obs) y data de prueba (30% de obs). Se ajustan dos modelos a la data: una regresión lineal simple y una regresión cúbica, es decir: $\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \beta_2 \mathbf{x}^2 + \beta_3 \mathbf{x}^3 + \epsilon$.
 - (a) Suponga que la verdadera relación entre \mathbf{x} e \mathbf{y} es lineal, es decir, $\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \epsilon$. Considere el MSE para la regresión lineal, y para la regresión cúbica utilizando data de entrenamiento. ¿Cómo se comparan estas dos cantidades? (1 punto)
 - (b) Responda (a) usando la data de prueba en lugar de la data de entrenamiento. (1 punto)