Laboratorio 1

Tópicos probabilísticos introductorios

Profesor: Felipe Tobar **Auxiliares**: Cristóbal Alcazar, Camilo Carvajal Reyes

Fecha de entrega: 8 de agosto 2023

Instrucciones: El siguiente laboratorio consiste en tres preguntas relacionadas con los tres tópicos probabilísticos vistos en clase. Se pedirá una entrega de un PDF de extensión máxima de 3 páginas en el caso de una pregunta teórica y una pieza de código comentada (ya sea en .ipynb o .py) para cada pregunta prácticas. Deberá cumplir lo siguiente.

1. Fije la semilla de aleatoriedad como su RUT.
2. Los comentarios en código deben ser concisos pero claros. No se evaluarán sub-preguntas donde solo exista código sin comentarios pertinentes.
3. El código debe ser ordenado y ejectuable. No se evaluarán notebooks o scripts que generen errores en su ejecución. Se aconseja resetear la kernel y correr todas las celdas de cero antes de entregar en el caso de un notebook.
4. Las partes teóricas deben ser legibles y explicar de manera clara lo pedido. No se evaluarán preguntas donde el mensaje no sea entendible para el equipo docente.

P1 - *Sampling* (40 %)

Nota: Si decide responder nuevamente, su puntuación se promediará con la entrega anterior; de lo contrario, solo se tendrá en cuenta la nota de la otra entrega para la P2.

(a) (1 pto.) Genere μ que distribuyan según una ley "semi normal" (*half-normal* en inglés), i.e., tal que $\mu = |X|$ (donde $|\cdot|$ denota el valor absoluto) con $X \sim \mathcal{N}(0, \sigma^2)$ para $\sigma^2 = 10$.

(b) (1 pto.) Genere n muestras $x_i \sim \mathcal{N}(\mu, 5)$.

(c) (1 pto.) Exprese $p(\mu|x_1, \ldots, x_n)$

(d) (1 pto.) Genere m muestras de $p(\mu|x_1, \ldots, x_n)$ usando Metropolis-Hastings. Puede seleccionar la distribución auxiliar q que le parezca adecuada.

(e) (1 pto.) Muestre los *samples* de Metropolis-Hastings para distintos valores de n y compare con el μ real.
Nota: seleccione muestras de la cadena considerando "thinning" y "burn in".

(f) (1 pto.) Repita todo el procedimiento anterior para un modelo de 3 componentes, es decir:

$$x_i \sim \sum_{i=1}^{3} \lambda_j \mathcal{N}(\mu_j, 5)\,.$$

Elija usted las distribuciones $p(\mu_1, \mu_2, \mu_3)$ y $\mu(\lambda_1, \lambda_2, \lambda_3)$.

P2 - Modelos gráficos (40 %)

Usted sospecha que tiene *Rinitis alérgica*. Como acostumbra a razonar de manera probabilista ante las situaciones, y dado que acaba de inscribir el ramo Modelos Generativos Profundos, se le ocurre plantear variables aleatorias para inferir la probabilidad de tener o no alergia. Se restringirá a las siguientes variables (que puede considerar binarias por el momento):

- Irritación en nariz y ojos (*Hint*: variable observada)
- Cantidad de polen en el ambiente
- Tener o no *Rinitis alérgica*
- Tener un familiar con *Rinitis alérgica* (suponga que no lo sabe)

(a) (2 ptos.) Plantee el problema anterior como un modelo gráfico tipo DAG (Grafo acíclico dirigido). Explicite claramente sus hipótesis, las variables y las dependencias entre ellas.

(b) (1 pto.) Entregue una expresión para la probabilidad de tener *Rinitis alérgica* dado que presenta irritación en nariz y ojos.

(c) (1.5 ptos.) Usted decide comprar un test de Frotis Nasal. Formule hipótesis e incorpore el resultado (positivo o negativo) del test en la probabilidad de tener o no *Rinitis alérgica*.

(d) (Bonus) Suponga ahora que la cantidad de polen en el ambiente es una variable continua Y_{polen} con ley "semi normal" (*half-normal* en inglés), i.e., tal que $Y_{polen} = |X|$ (donde $|\cdot|$ denota el valor absoluto) con $X \sim \mathcal{N}(0, \sigma^2)$ para algún σ dado. Explique cómo distribuye la cantidad de polen en el ambiente ¿Cómo cambia respecto al modelo anterior?

Usted concluye y asume en adelante que tiene alergia. A usted le interesa modelar la cantidad de polen en el ambiente para predecir altos niveles y prepararse ante el malestar. Sin embargo un detector de polen supera su prespuesto.

(e) (1.5 ptos.) Plantee el problema de predecir la cantidad de polen cada 30 minutos. Usted usará para esto su nivel de irritación en nariz y ojos como única medición. Explicite sus variables, y asuma que el futuro es independiente del pasado, dado el presente.

P3 - Introducción al Transporte Óptimo (40 %)

Considere que el transporte óptimo entre dos distribuciones discretas $\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}$ y $\beta = \sum_{j=1}^{m} b_j \delta_{y_j}$ está dado por

$$P^* \in \arg \min_{P \in U(a,b)} \sum_{i,j} P_{i,j} C_{i,j} \,,$$

donde $U(a,b) := \{\ P \in \mathbb{R}_+^{n \times m}, \forall i, \sum_j P_{i,j} = a_i, \forall j, \sum_i P_{i,j} = b_j\}$ y $C_{i,j} = \|x_i - x_j\|^2$.

(a) (2 ptos.) Usando algún *solver* de programación lineal, implemente un método que resuelva el problema.

Sean $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ y $x_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ dos distribuciones Gaussianas en $\mathbb{R}^2$. Para el caso $d = 1$, el transporte óptimo está dado por la siguiente expresión cerrada:

$$x \to \mu_2 + A(x - \mu_1) \,, A = \Sigma_1^{-\frac{1}{2}} (\Sigma_1^{\frac{1}{2}} \Sigma_2 \Sigma_1^{\frac{1}{2}})^{\frac{1}{2}} \Sigma_1^{-\frac{1}{2}} \,,$$

con lo cual la distancia de Wasserstein es:

$$W_2(\mathcal{N}_1(\mu_1, \Sigma_1), \mathcal{N}(\mu_2, \Sigma_2)) = \|\mu_1 - \mu_2\|_2^2 + Tr(\Sigma_1 + \Sigma_1 + \Sigma_2 - 2(\Sigma_1^{\frac{1}{2}} \Sigma_2 \Sigma_1^{\frac{1}{2}})) \,.$$

(b) (Bonus) Comente la intuición detrás del la expresión cerrada del transporte óptimo entre dos gaussianas.

(c) (1 pto.) Samplee de las distribuciones para parámetros de su elección y resuelva el método anterior para $d = 1$. Compare con el valor teórico.

(d) (1 pto.) Samplee y grafique dos Gaussianas para el caso $d = 2$. Use el método para calcular el transporte óptimo y grafique las asignaciones que induce.

Para dos distribuciones α y β, y $t \in [0, 1]$ podemos definir el baricentro de Wasserstein como:

$$\mu_t = \arg \min_{\mu} (1 - t) W_2(\alpha, \mu)^2 + t W_2(\beta, \mu)^2 \,,$$

que puede entenderse como la generalización del baricentro euclideano para distribuciones. Una vez que tenemos acceso al transporte óptimo P^*, la interpolación en cuestión está dada por

$$\mu_t = \sum_{i,j} P^*_{i,j} \delta_{(1-t)x_i + t y_j} \,.$$

(e) (2 ptos.) Para una cantidad adecuada de pasos $t_0 = 0 < t_1 < \cdots < t_k = 1$, compute y grafique las distribuciones empíricas para el baricentro $\mu_{t_i}, i = 1, \ldots, k$.