BOURBAKI

COLEGIO DE MATEMÁTICAS

BOURBAKI
COLEGIO DE MATEMÁTICAS

Índice

01 Introducción

Bienvenidos a nuestro curso de Matemáticas Avanzadas para la Ciencia de Datos, nuestro curso tiene tres módulos dedicados a estudiar las ideas matemáticas más útiles para comprender los algoritmos y modelos matemáticos más comúnes en Ciencia de Datos. Los tres módulos son los siguientes

- Probabilidad y Estadística
- Álgebra Lineal
- Optimización y cálculo diferencial

Todos los módulos tienen una duración de 8 semanas. El curso está acompañado de ejercicios y tareas en Python para practicar y reforzar los conocimientos aprendidos así como las implementaciones en bases de datos de los algoritmos estudiados. Pueden consultar el repositorio de esta semana en este link.

La estructura de cada una de las semanas es la siguiente:

1. Treinta minutos dedicados a estudiar un artículo de referencia que motivará los conceptos matemáticos de esta semana.
2. Dos horas dedicadas a estudiar el tema de la semana y algunos ejercicios.

3. Una hora y media dedicada a practicar lo aprendido utilizando Python.

El primer módulo de probabilidad consta de los siguientes temas:

1. Kolmogorov, independencia y condicionamiento
2. Variables aleatorias discretas y sus momentos
3. Ley de los grandes números y máxima verosimilitud
4. El teorema límite central y los intervalos de confianza
5. Tests estadísticos
6. Método de monte carlo para cadenas de markov
7. Inferencia bayesiana
8. Redes bayesianas

02 Lectura de referencia: la simulación de markov y la criptografía

La simulación de monte carlo se ha utilizado ampliamente en matemáticas aplicadas para resolver problemas de muy distintas áreas. En el texto recomendado de esta semana ([1]) el mago y matemático profesional Persi Diaconis reflexiona sobre la revolución que ha supuesto para muchas áreas del conocimiento el uso de un método de monte carlo para cadenas de markov adecuado para problemas de muy diversa índole.

Una de las aplicaciones que discutiremos con los alumnos esta semana es cómo utilizar el método de monte carlo para descifrar mensajes encriptados. En la referencia se mencionan algunos textos y en general la estrategia pasa por encontrar una medida de probabilidad adecuada con soporte en la solución deseada para luego dejar a la simulación encontrar los parámetros.

03 Cadenas de Markov

Definition 00.1.

- Un proceso estocástico discreto es una familia $(X_t)_{t\in\mathbb{N}}$ de variables aleatorias $X_t : \Omega \to \mathbb{R}^d$.

- Un proceso estocástico $(X_t)_{t\in\mathbb{N}}$ es una cadena de markov si para todo $t \in \mathbb{N}$ y todos $(r_1, \dots, r_t, r_{t+1}) \in \mathbb{R}^{d(t+1)}$ se tiene:

$$\mathbb{P}\left(X_{t+1} = r_{t+1} | X_1 = r_1, \dots, X_t = r_t\right) = \mathbb{P}\left(X_{t+1} = r_{t+1} | X_t = r_t\right)$$

- Una cadena de markov $(X_t)_{t\in\mathbb{N}}$ es homogénea si para todo $\left(t, t'\right) \in \mathbb{N}^2$ y todos $(r, s) \in \mathbb{R}^{d\cdot 2}$

$$\mathbb{P}\left(X_{t+1} = r | X_t = s\right) = \mathbb{P}\left(X_{t'+1} = r | X_{t'} = s\right)$$

- Asociada a cada cadena de markov homogénea $(X_t)_{t\in\mathbb{N}}$, $X_t : \Omega \to \mathbb{R}^d$ existe matriz M tal que para cada $r, s \in \mathbb{R}^d$, $t, i \geq 0$

$$M_{(r,s)} = \mathbb{P}\left(X_{t+1} = r | X_t = s\right) = \mathbb{P}\left(X_{t+i} = r | X_{t+i-1} = s\right)$$

Remark 00.1. *Notemos que la matriz M asociada a una cadena de markov homogénea en un principio no tiene por qué ser finita, sin embargo al estar suponiendo que nuestros espacios de probabilidad son todos finitos (lo cuál implica que nuestras cadenas de markov son discretas) entonces solo una can-*

tidad de entradas $M_{r,s}$ de la matriz serán distintas a cero.

Exercise 00.2. *Sea $(X_t)_t$ un proceso de variables aleatorias independientes e idénticamente distribuidas. Demostrar que ellas son una cadena de Markov.*

Exercise 00.3. *Supongamos que Y_t es una familia de variables aleatorias independientes e identicamente distribuidas. Entonces las variables aleatorias $X_t = Y_1 + Y_2 + \ldots + Y_t$ son una cadena de markov.*

Exercise 00.4. *Sea $G = (V, E)$ un grafo finito, para un $v \in V$ denotamos por E_{v_i} a los elementos $v_j \in V$ tales que $(v, w) \in E$. Definimos la siguiente matriz de markov:*

- $M_G\left(v_i, v_j\right) = \frac{1}{|E_{v_i}|}$ *si* $\left(v_i, v_j\right) \in E$
- $M_G\left(v_i, v_j\right) = 0$ *si* $\left(v_i, v_j\right) \notin E$

Remark 00.5. *Una vez que hemos fijado Ω y la imagen posible de las variables aleatorias, existe una equivalencia entre cadenas de markov discretas (y homogéneas) con sus matrices asociadas, por su puesto que debemos pedir que la suma de todas las entradas en una columna sumen uno, esta construcción lleva el nombre de cadena de markov canónica.*

Matrices estocásticas

Algunas veces será más sencillo representar una cadena de markov mediante su matriz $M_{s,r}$ asociada:

Definition 01.1. Dado un conjunto finito S llamado conjunto de estados, una matriz estocástica es una función $M : S \times S \to \mathbb{R}$ tal que:

- Para todos $s, r \in S$, $M(s, r) \in [0, 1]$
- Para cualquier $s \in S$, $\sum_{r \in S} M(s, r) = 1$

Proposition 01.1. *Dada una matriz estocástica M sobre un conjunto finito S y $t \in \mathbb{N}$ es posible construir otra matriz estocástica M_t de manera inductiva utilizando la siguiente fórmula:*

1. $M_1 = M$
2. *Si* $t \geq 2$, $M_t(s, s') = \sum_{r \in S} M_{t-1}(s, r) M(r, s')$

Exercise 01.2. *Encontrar una matriz estocástica asociada a los procesos estocásticos de los ejercicios 00.2 y 00.3.*

La ruina del apostador

Sean N un número entero positivo que en esta sección corresponde a la cantidad de pesos que un apostador desea ganar. Supongamos que este apostador comienza a jugar con una cantidad $0 < c < N - 1$.

El juego consiste en lanzar una moneda, y de acuerdo al resultado en esa moneda, el apostador gana un peso o pierde un peso. Si suponemos que los lanzamientos son independientes e idénticamente distribuidos entonces podemos pensar en la sucesión de los lanzamientos como un conjunto de variables aleatorias X_i tales que

$$\mathbb{P}(X_i = 1) = p, \mathbb{P}(X_i = -1) = 1 - p$$

Utilizando este conjunto de lanzamientos podemos definir el proceso estocástico que modela los estados en los que se encuentra el apostador, Y_i es un conjunto de variables aleatorias tales que:

$$Y_i = c + X_1 + X_2 + \ldots + X_i$$

Si suponemos que además se satisface la siguiente regla: cuando el apostador no tenga más dinero o tenga N pesos, el juego se acaba, en este caso podemos notar que este proceso es una cadena de Markov cuyo conjunto de estados es $\{0, 1, 2, \ldots, N\}$ y satisface lo siguiente para $i \neq 0, N$ y $j \geq 0$:

$$\mathbb{P}\left(Y_{j+1} = Y_j + 1 | Y_j = i\right) = p, \mathbb{P}\left(Y_{j+1} = Y_j - 1 | Y_j = i\right) = 1 - p$$

Para los otros posibles valores de i las probabilidades de trancisión son las siguientes:

$$\mathbb{P}\left(Y_{j+1} = 0 | Y_j = 0\right) = 1, \mathbb{P}\left(Y_{j+1} = N | Y_j = N\right) = 1$$

Proposition 02.1. *El proceso Y_i define una cadena de Markov.*

Exercise 02.2. *Demostrarlo.*

Definition 02.1. Definamos la variable aleatoria $\tau_c = \underset{n\geq 0}{argmin}(Y_n \in \{0, N)$ y $p_{c,N} = \mathbb{P}\left(Y_{\tau_c} = N\right)$.

Exercise 02.3. *Explicar con palabras la interpretación de* $\tau_c, p_{c,N}$.

Exercise 02.4. *Demostrar que* $p_{0,N} = 0, p_{N,N} = 1$.

El teorema de la ruina del apostador dice lo siguiente:

Theorem 02.5. *Supongamos la cadena de markov* Y_i *definida anteriormente.*

- *Si* $p = \frac{1}{2}$ *entonces* $p_{c,N} = \frac{c}{N}$
- *Si* $p \neq \frac{1}{2}$ *entonces* $p_{c,N} = \frac{1-\frac{1-p}{p}}{1-\left(\frac{1-p}{p}\right)^N}$

Teoremas de ergodicidad y medidas estacionarias

Definition 03.1. Sea $X = (X_t)_t$ una cadena de markov discreta (sobre Ω digamos), aperiódica e irreducible con matriz de trancisión M. Diremos que una medida de probabilidad $\mathbb{P}_\mu$ en el conjunto Ω es una medida estacionaria de X cuando satisfaga la ecuación Chapman-Kolmogorov i.e. para cualquier $\omega \in \Omega$:

$$M\mathbb{P}_\mu := \sum_{\omega' \in \Omega} \mathbb{P}_\mu(\omega)\, M\left(\omega, \omega'\right) = \mathbb{P}_\mu(\omega)$$

Remark 03.1. *Por el momento no hemos enunciado ni existencia ni unicidad de la distribución* μ*, sin embargo vale la pena pensar en la medida* $\mathbb{P}_\mu$ *como un comportamiento límite de la cadena* X.

Exercise 03.2. *Sea* $G = (V, E)$ *un grafo y* $(X_t)_t$ *su cadena de markov asociada como en el ejemplo 00.4. Definimos* $\mathbb{P}_\mu(v) := \frac{1}{|E_v||V|}$*, dejamos como ejercicio al lector verificar que efectivamente* μ *es una medida invariante. La interpretación como un comportamiento límite de esta medida de probabilidad es gracias a la cantidad sin normalizar:* $|V|\mathbb{P}_\mu(v) = \frac{1}{E_v}$ *que corresponde a "con qué probabilidad estaría yo en* v*?"*

Exercise 03.3. *Supongamos que una cadena de markov* $(X_t)_t$ *satisface lo siguiente:*

- $\mathbb{P}_{X_0} = \mathbb{P}_\mu$
- $\mathbb{P}_\mu$ *es una medida invariante de la cadena de markov*

Entonces $\mathbb{P}_{X_n} = \mathbb{P}_\mu$ *para todo* $n \geq 0$.

El siguiente teorema es tiene la misma importancia que la ley de los grandes números, esta vez para cadenas de markov en lugar de variables independientes e idénticamente distribuidas:

Theorem 03.4. *(Teoremas de ergodicidad para cadenas de markov) Sea* $(X_t)_t$ *una cadena de markov discreta (sobre* Ω *digamos), aperiódica e irreducible con matriz de trancisión* M*. Entonces existe un conjunto* $N \subseteq \Omega$ *con* $\mathbb{P}_\mu(N) = 0$*:*

1. *La medida invariante μ existe y es única.*

2. *(Ley fuerte los grandes números para cadenas de markov)* $\lim_{n\to\infty} \frac{1}{n} \sum_{i \leq n-1} \mathbf{1}_{X_i = \omega} = \mathbb{P}_{\mu}(\omega)$ *siempre que* $\omega \notin N$.

3. *(Teorema límite central para cadenas de markov) Si definimos* $M_n = \left[\sqrt{n}\left(\frac{1}{n} \sum_{i \leq n-1} \mathbf{1}_{X_i = \omega}\right) - \mathbb{P}_{\mu}\right]$ *entonces*

$$\lim_{n\to\infty} \mathbb{P}_{M_n}(a,b) = \mathbb{P}_{Gauss,0,\sigma^2}(a,b)$$

   Para algún σ^2.

4. $\lim_{n\to\infty} \mathbb{P}_{X_n}(a,b) \to \mathbb{P}_{\mu}(a,b)$

Corollary 03.5. *El modelo de Page Rank tiene solución.*

Remark 03.6. *Es posible demostrar que* $\mathbb{P}_{\mu}(\omega) = \frac{1}{\mathbb{E}[T_{\omega}]}$ *para cierta variable aleatoria* T_{ω} *que no definiremos.*

Exercise 03.7. *Demostrar que el teorema anterior efectivamente generaliza la ley de los grandes números.*

Bibliografía

[1] Persi Diaconis. The markov chain monte carlo revolution. Bulletin of the American Mathematical Society, 46(2):179–205, 2009.

BOURBAKI

COLEGIO DE MATEMÁTICAS

escuela-bourbaki.com