Lista de ejercicios 3

Curso: Tópicos de Investigación : Machine Learning CM-072

Lecturas Importantes

1. Interesante guia de data mining: `http://guidetodatamining.com/`.

Probabilidad

1. Un dado es lanzado n veces. Muestra que la probabilidad de que haya un número par de `seis` $\frac{1}{2}[1 + (\frac{2}{3})^n]$. Para este caso 0 es un número par.

2. Sea $A_r, r \geq 1$, los eventos tal que $P(A_r) = 1$ para todo r. Muestra que $P(\cap_{r=1}^{\infty} A_r) = 1$.

3. Consideremos $\Omega = \{1, 2, \ldots, p\}$ donde p es primo. Sea F el conjunto de todos los subconjuntos de Ω y $P(A) = |A|/p$ para todo $A \in F$. Muestra que, si A y B son eventos independientes, entonces al menos unos de los conjuntos A y B es $\emptyset$ o Ω.

4. Si m estudiantes nacidos en en 1991 están asistiendo a una conferencia, muestra que la probabilidad de que al menos dos de ellos tengan el mismo cumpleaños es $p = 1 - (365)!/(365 - m)!365^m$. Muestra que $p > 1/2$ cuando $m = 23$.

5. Para que valores de c y α es la función p definida por

$$p(k) = \begin{cases} ck^{\alpha} & \text{para } \; k = 1, 2, \ldots \\ 0 & \text{en otros casos} \end{cases}$$

   es una función de masa de probabilidad.

6. Si X tiene una distribución geométrica con parámetro p, muestra que

$$P(X > m + n | X > m) = P(X > n)$$

   para $m, n = 0, 1, 2, \ldots$.

7. Sea X y Y una variable aleatoria discreta, cada teniendo una función de masa dado por

$$P(X = k) = P(Y = k) = pq^k \;\text{ para }\; k = 0, 1, 2, \ldots$$

   donde $0 < p = 1 - q < 1$. Muestra que

$$P(X = k | X + Y = n) = \frac{1}{n+1} \;\text{ para }\; k = 0, 1, 2, \ldots, n.$$

8. Sea $X_1, X_2, \ldots$ variables aleatoria independientes, identicamente distribuidas y sea $S_n = X_1 + X_2 + \cdots, +X_n$. Muestra que $E(S_m/S_n) = m/n$ si $m \leq n$ y $E(S_m/S_n) = 1 + (m - n)\mu E(1/S_n)$ si $m > n$, donde $\mu = E(X_1)$.

9. Sean F y G funciones de distribución y sea la `métrica de Levi`

$$d_L = \inf\{\epsilon > 0 : G(x-\epsilon) - \epsilon \leq F(X) \leq G(x+\epsilon) + \epsilon \text{ para todo } x\}$$

Muestra que d_L es efecto una métrica sobre el espacio de las funciones de distribución.

10. Sea $X_r, 1 \leq r \leq n$, variables aleatorias independientes, que son simétricas alrededor del 0; esto es, X_r y X_r tienen la misma distribución. Muestra que, para todo x, $P(S_n \geq x) = P(S_n \leq -x)$, donde $S_n = \sum_{r=1}^{n} X_r$.

11. Sea $G(V,E)$ un grafo finito. Para un conjunto de vértices y una arista $e \in E$, definimos la función indicador

$$I_W(e) = \begin{cases} 1 & \text{si } e \text{ conecta } W \text{ y } W^c \end{cases}$$

Sea $N_W = \sum_{e \in E} I_W(e)$. Muestra que existe un $W \subseteq N$ tal que $N_W \geq \frac{1}{2}|E|$.

12. Sea X con una función generadora de probabilidad $G_X(s)$ y sea $u_n = P(X > n)$. Muestra que la función generadora $U(s)$ de la secuencia $u_0, u_1, \ldots$ satisface

$$(1-s)U(s) = 1 - G_X(s)$$

siempre que la serie definida, de esa función generadora converge.

13. Sean X e Y variables aleatorias independientes, con una distribución de Poisson con parámetros μ y λ respectivamente. Prueba que $X+Y$ tiene una distribución de Poisson y que $var(X+Y) = var(X) + var(Y)$. Encuentra la probabilidad condicional $P(X = k|X + y = n)$ para $0 \leq k \leq n$ y así muestra que la esperanza condicional de X dado $X+Y = n$, esto es

$$E(X|X+Y=n) = \sum_{k=0}^{\infty} kP(X=k|X+Y=n).$$

es $n\lambda/(\lambda+\mu)$.

14. Una moneda muestra cara con probabilidad p. Sea X_n el número de de lanzamientos requeridos para obtener n caras consecutivas. Muestra que $E(X_n) = \sum_{k=1}^{n} p^{-k}$.

15. Si $var(X) = 0$, entonces, existe un $a \in \mathbb{R}$ tal que $P(X = a) = 1$.

16. Sean X, Y variables aleatorias independientes con una función de distribución común F y una función densidad f. Muestra que $V = \max X, Y$ tiene una función de distribución $P(V \leq x) = F(x)^2$ y función densidad $f_V(x) = 2f(x)F(x), x \in \mathbb{R}$. Encuentra la función densidad de $U = \min\{X, Y\}$

17. Sea X una variable aleatoria con media μ y una función de distribución continua F. Muestra que

$$\int_{-\infty}^{a} F(x)dx = \int_{a}^{\infty} [1 - F(x)]dx$$

18. Muestra que para la densidad normal estándar $\phi(x)$, muestra que $\phi^{'}(x) + x\phi(x) = 0$.

19. Sea $\{X_r : r \geq 1\}$ variables aleatorias indepedientes, uniformemente distribuidas en $[0,1]$. Sea $0 < x < 1$ y definimos

$$N = \min\{n \geq 1 : X_1 + X_2 + \cdots + X_n > x\}$$

Muestra que $P(N > n) = x^n/n!$ y así encuentra la media y la varianza de N.

20. Sea X una variable aleatoria con una distribución binomial con parámetros n y p, muestra que

$$E\Big(\frac{1}{1+X}\Big) = \frac{1-(1-p)^{n+1}}{(n+1)p}$$

Encuentra el límite de esta expresión cuando $n \to \infty$ y $p \to \infty$.

21. La variable aleatoria X tiene una función densidad proporcional a $g(x)$, donde g es una función satisfaciendo

$$g(x) = \begin{cases} |x|^{-n} & \text{si} |x| \geq 1 \\ 0 & \text{en otros casos} \end{cases}$$

y $n \geq 2$ es un entero. Encuentra la función de densidad de X y determina los valores de n para el cual la media y la varianza de X existe.

22. Si X es una variable aleatoria que toma los valores no negativos, muestra que

$$E(X) = \int_0^\infty [1 - F_X(x)]dx$$

siempre que la integral exista.

23. Es la función G, definida por

$$G(X,Y) = \begin{cases} 1 & \text{if } x+y \leq 0, \\ 0 & \text{en otros casos} \end{cases}$$

una función de distribución conjunta de algún par de variables aleatorias.

24. Sea X y Y variables aleatorias independientes, X teniendo la distribución normal con media 0 y la varianza 1 y Y teniendo la distribución χ^2 con n grados de libertad. Muestra que

$$T = \frac{X}{\sqrt{Y/n}}$$

tiene una función densidad

$$f(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \Big(1 + \frac{t^2}{n}\Big)^{-\frac{1}{2}(n+1)} \quad \text{para } t \in \mathbb{R}.$$

25. Sea X una función generadora de momentos $M(t)$.

- Muestra que $M(t)M(-t)$ es la función generadora de momentos de $X - Y$, donde Y es independiente de X , pero que tiene la misma distribución.
- De manera similar, describe las variables aleatorias que tienen funciones generadoras de momentos

$$\frac{1}{2-M(t)}, \qquad \int_0^\infty M(ut)e^{-u}du.$$

Álgebra Lineal

1. Calcula la dimensión de cada uno de los conjuntos

- $col \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

- span$\{(1,1,1),(1,-1,1),(-1,1,1),(1,1,-1)\}$.
- span$\{(2,7,9),(3,5,1),(0,1,0)\}$.
- $col\begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$

2. Supongamos que $A, B \in \mathbb{R}^{n\times n}$ y $a, b \in \mathbb{R}$. Encuentra un sistema de ecuaciones lineales que para un x minimize la función $\|Ax-a\|^2 + \|Bx-b\|^2$.

3. Supongamos que $A \in \mathbb{R}^{n\times n}$ es simétrica y definimos el cociente de Rayleigh como

$$R(x) = \frac{x^T Ax}{\|x\|^2}$$

   Muestra que los minimizadores de $R(x)$ sujetos a $x \neq 0$ son los autovectores de A.

4. Muestra que toma $O(n^2)$ tiempo para encontrar el producto AB de dos matrices $A, B \in \mathbb{R}^{n\times n}$.

5. Sea $p(X) : \mathbb{R}^n \to [0,1]$, tal que $p(x) \geq 0$ y para todo $x \in \mathbb{R}^n$ y

$$\int_{\mathbb{R}^n} p(x)dx = 1$$

   Un tipo importante de esta función es la `distribución normal` que tiene la forma

$$G_{\Sigma,\mu}(x) = e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

   La matriz covarianza $\Sigma \in \mathbb{R}^{n\times n}$ y la media $\mu \in \mathbb{R}^n$, determina la fórmula de la distribución normal. Suponiendo que x^* es un máximo local de $p(x)$. Sugiere una aproximación de $p(x)$ en una vecindad de x^*.

6. Factorizar la matriz A como un producto $A = LU$:

$$\begin{pmatrix}1 & 2 & 7\\ 3 & 5 & -1\\ 6 & 1 & 4\end{pmatrix}$$

7. Supongamos que tenemos una matriz cuadrada $M \in \mathbb{R}^{n\times n}$ es escrita en forma de bloques

$$M = \begin{pmatrix}A & B\\ C & D\end{pmatrix}$$

   donde $A \in \mathbb{R}^{k\times k}$.

   - Muestra que podemos descomponer M como el producto

$$M = \begin{pmatrix}I & O\\ CA^{-1} & I\end{pmatrix}\begin{pmatrix}A & O\\ 0 & D-CA^{-1}B\end{pmatrix}\begin{pmatrix}I & A^{-1}B\\ O & I\end{pmatrix}$$

     donde I es la matriz identidad de tamaño apropiado.
   - Supongamos que descomponemos $A = L_1U_1$ y $D - CA^{-1}B = L_2U_2$. Muestra como construir una factorización LU de M dados esas matrices adicionales.
   - Usa esta estructura para definir un algoritmo recursivo para la factorización LU, se puede asumir que $n = 2^l$ para algún $l > 0$.

8. Muestra que una matriz invertible $A \in \mathbb{R}^{n\times n}$ con $a_{11} = 0$ no puede tener una factorización $A = LU$ para una matriz triangular L y una matriz triangular superior U.

9. Escribe una función para resolver el sistema $Ax = b$, usando la descomposición SVD. La función debe tomar A y b como entrada y debe retornar x. La función debe incluir lo siguiente

   - Verificar si A es inversible, si no es así debe retornar un mensaje de error.
   - Invertir A usando SVD y resolver.
   - Retornar x

10. Implementa el algoritmo de Gedanken: Dada una matriz A de orden $m \times n$, encuentra vectores $v_1, \ldots, v_{\text{rank A}}$ tal que, para $k = 1, 2, \ldots$ rank (A), el subespacio k-dimensión V que minimiza

$$\sum_i (\text{distancia de la fila } i \text{ de } A \text{ hasta } V_k)^2$$

    es el `Span` $\{v_1, \ldots, v_k\}$.

11. Supongamos que $A \in \mathbb{R}^{m\times n}$ tal que $m < n$. Muestra que tomando $x = A^T(AA^T)^{-1}b$ resuelve el problema de optimización:

$$\begin{aligned} &\min_x \|x\| \\ &\text{sujeto } \; Ax = b \end{aligned}$$

    Muestra que tomando $\alpha \to 0$ en el sistema de Tikhonov recupera el valor elegido de x.

    `Revisar: Numerical Algorithms de Justin Solomon.`

12. Da un ejemplo de matriz sparse cuya inversa es densa.

13. Supongamos que $A \in \mathbb{R}^{n\times n}$ admite una factorización de Cholesky $A = LL^T$

    - Muestra que A debe ser semidefinida positiva.
    - Use esta observación para sugerir un algoritmo para verificar si una matriz es semidefinida positiva.

14. Muestra como las técnicas lineales puede ser usadas para resolver los siguientes problemas de optimización para $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{K\times n}, c \in \mathbb{R}^k$

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \|Ax\| \\ &\text{sujeto } \; Bx = c \end{aligned}$$

15. Una versión general de la descomposición de Cholesky, que no requiere el cálculo de las raices es la `descomposición LLT`

    - Supongamos que $A \in \mathbb{R}^{n\times n}$ es simétrica y admite una factorización LU (sin pivot). Muestra que A puede ser factorizada como $A = LDL^T$, donde D es diagonal y L es una matriz triangular.
    - Modifica la construcción de la descomposición de Cholesky para mostrar que una matriz simétrica, definida positiva A puede ser factorizada como $A = LDL^T$ sin usar alguna operación de raices cuadradas.