Homework#0

Zhaoxing Deng 005024802

January 18, 2018

1. $\frac{\partial y}{\partial x} = \sin(z)\exp(-x) + x\sin(z)\exp(-x)(-1) = \sin(z)(1-x)\exp(-x)$

2. Linear Algebra

   (a) $\mathbf{y}^T\mathbf{z} = \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2 + 9 = 11$

   (b) $\mathbf{xy} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$

   (c) $\left(\begin{array}{cc:cc} 2 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right) \rightarrow \left(\begin{array}{cc:cc} 1 & 2 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array}\right) \rightarrow \left(\begin{array}{cc:cc} 1 & 0 & \frac{2}{3} & -2 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array}\right)$
   $inv(\mathbf{x}) = \begin{pmatrix} \frac{2}{3} & -2 \\ -\frac{1}{2} & 1 \end{pmatrix}$

   (d) $\begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 4 \\ 0 & 1 \end{pmatrix}$
   $\text{rank}(\mathbf{x}) = 2$

3. Probability and Statistics

   (a) $\bar{X} = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5) = \frac{3}{5}$

   (b) $\sigma^2 = \frac{1}{5-1}\sum_{i=1}^{5}\left(X_i - \bar{X}\right)^2 = 0.3$

   (c) $P(S) = P(X_1 = 1)\,P(X_2 = 1)\,P(X_3 = 0)\,P(X_4 = 1)\,P(X_5 = 0) = 0.3125$

   (d) Suppose $P(X_i = 1) = p$
   $P(S) = P(X_1 = 1)\,P(X_2 = 1)\,P(X_3 = 0)\,P(X_4 = 1)\,P(X_5 = 0) = p^3(1-p)^2$
   Let $\frac{\partial P(S)}{\partial p} = 0$, then we can get $p^2(5p-3)(p-1) = 0$
   $\because 0 < p < 1 \therefore p = \frac{3}{5}$

   (e) $P(X = T|Y = b) = \frac{P(X=T,Y=b)}{P(Y=b)} = \frac{P(X=T,Y=b)}{P(X=T,Y=b)+P(X=F,Y=b)} = 0.4$

4. Probability axioms

   (a) false
   (b) true
   (c) false
   (d) false
   (e) true

5. Discrete and Continuous Distributions

   (a) (v)
   (b) (iv)
   (c) (ii)
   (d) (i)
   (e) (iii)

6. Mean and Variance

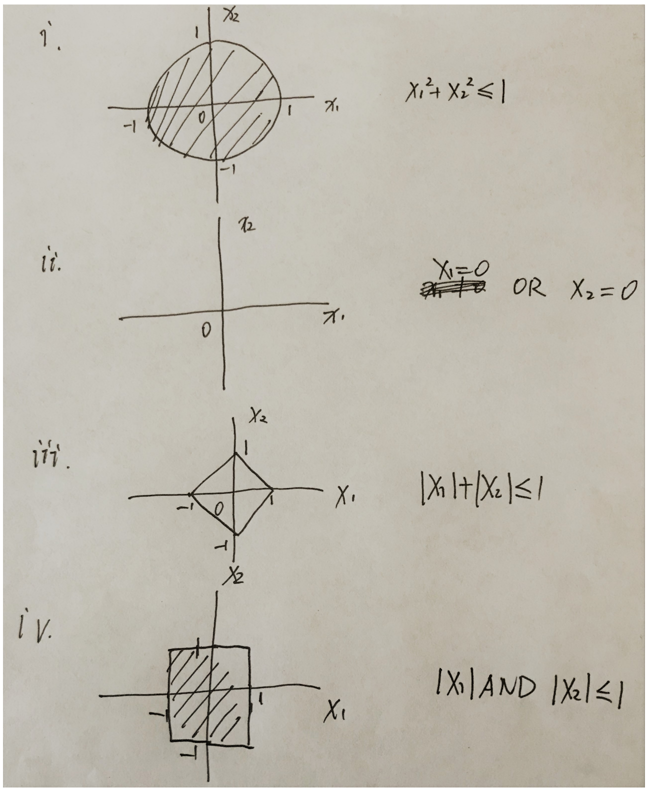
   (a) The PMF of$Bernoulli(p)$ is
   $P(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$
   $E[X] = 1 \cdot p + 0 \cdot (1-p) = p$
   $Var[X] = E[X^2] - (E[X])^2 = p(1-p)$
   (b) $Var[2X] = 2^2 Var[X] = 4\sigma^2$
   $Var[X+2] = Var[X] + Var[2] + Cov[X,2] = \sigma^2$

7. Algorithms

   (a) Big-O notation
      i. $f(n) = O(g(n)),\ g(n) = O(f(n))$
      ii. $f(n) = O(g(n))$
      iii. $f(n) = O(g(n)),\ g(n) = O(f(n))$
   (b) Our goal is to find the index i such that the ith element in the array is 0 and the (i+1)th element is +1.
   We can use Divide and Conquer method. Each time, we set i equals the middle of array's length.
   If ith element is 0, turn to the left half of the array. If ith element is +1, turn to the right half of the array.

8. Probability and Random Variables

   (a) Suppose X, Y are discrete random variables
   $E[X]\cdot E[Y] = \left(\sum_x xP(X=x)\right)\left(\sum_y yP(Y=y)\right) = \sum_{x,y} xyP(X=x)P(Y=y)$
   $= \sum_z z\left(\sum_{x,y,xy=z} P(X=x)P(Y=y)\right) = \sum_z zP(XY=z)$
   $= E[XY]$
   (b) Law of Large Numbers and Central Limit Theorem
      i. Let$\begin{cases} X_i = 1 & \text{when 3 shows up in the ith toss} \\ X_i = 0 & \text{else} \end{cases}$
      Y denote the total number of times when 3 shows up
      $Y = \sum_{i=1}^{n} X_i$
      $\mu = E[X_i] = 1 \cdot p + 0 \cdot (1-p) = \frac{1}{6}$
      According to LLN, $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$, when $n \to \inf$
      $\therefore Y = n \cdot \bar{X_n} \approx n \cdot \mu =$6000 $\times \frac{1}{6} = 1000$



ii. $\mu = E[X_i] = \frac{1}{2}$
$\sigma^2 = Var[X_i] = E[X_i^2] - (E[X_i])^2 = \frac{1}{4}$
According to CLT, $\sqrt{n}\left(\bar{X}-\mu\right) \underrightarrow{n\to\inf} N\left(0,\sigma^2\right)$
$\therefore \sqrt{n}\left(\bar{X}-\frac{1}{2}\right) \underrightarrow{n\to\inf} N\left(0,\frac{1}{4}\right)$

9. Linear Algebra

(a)

(b) Matrix Decompositions

i. Given a square matrix$A \in \mathbb{R}^{n\times n}$, $\lambda \in C$ is an eigenvalue of $A$and$\mathbf{x} \in \mathbb{C}^n$ is the corresponding eigenvector if $A\mathbf{x} = \lambda\mathbf{x}$,$\mathbf{x} \neq \mathbf{0}$

ii. Let det$\begin{pmatrix} \lambda-2 & -1 \\ -1 & \lambda-2 \end{pmatrix} = 0$, we get$(\lambda-2)^2 - 1 = 0$
$\therefore \lambda = 1$ or 3
When $\lambda = 1$, $\mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
When $\lambda = 3$, $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

iii. Proof by induction
Let $\lambda_i$and $\mathbf{x}_i$,$i = 1, 2, ..., n$, be eigenvalues and corresponding eigenvectors of A
When k $= 1$, $A\mathbf{x}_i = \lambda_i^1\mathbf{x}_i$by definition, so in this case the conclusion is true.

Suppose when k = N, λ_1^N,λ_2^N,...,$\lambda_n^N$are the eigenvalues of matrix A^N, $\mathbf{x}_1$,$\mathbf{x}_2$,...,$\mathbf{x}_n$are corresponding eigenvectors.
When k = N+1, $A^{N+1}\mathbf{x}_i = A\left(A^N\mathbf{x}_i\right) = A\left(\lambda_i^N\mathbf{x}_i\right) = \lambda_i^N A\mathbf{x}_i = \lambda_i^{N+1}\mathbf{x}_i$, the conclusion is also true for k =N+1.

(c) Vector and Matrix Calculus

i. Let $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} \in \mathbb{R}^n$, $\mathbf{a} = \begin{pmatrix} a^1 \\ a^2 \\ ... \\ a^n \end{pmatrix} \in \mathbb{R}^n$,$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = \sum a_i x_i$

$$\nabla_{\mathbf{x}} f\left(\mathbf{x}\right) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ ... \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_n \end{pmatrix} = \mathbf{a}$$

ii. Let $g(\mathbf{x}) = \mathbf{x}^T A\boldsymbol{x}$=$\sum_i\sum_j A_{ij}x_ix_j$, A is symmetric, so $A_{ij} = A_{ji}$

$$\nabla_{\mathbf{x}} g\left(\mathbf{x}\right) = \begin{pmatrix} \frac{\partial g(\mathbf{x})}{\partial x_1} \\ \frac{\partial g(\mathbf{x})}{\partial x_2} \\ ... \\ \frac{\partial g(\mathbf{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \sum A_{1i}x_i + \sum A_{i1}x_i \\ \sum A_{2i}x_i + \sum A_{i2}x_i \\ ... \\ \sum A_{ni}x_i + \sum A_{in}x_i \end{pmatrix} = \begin{pmatrix} 2\sum A_{1i}x_i \\ 2\sum A_{2i}x_i \\ ... \\ 2\sum A_{ni}x_i \end{pmatrix} = 2A\mathbf{x}$$

$$\nabla^2_{\mathbf{x}} g\left(\mathbf{x}\right) = \begin{pmatrix} \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_2} & ... & \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 g(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 g(\mathbf{x})}{\partial x_2 \partial x_2} & ... & \frac{\partial^2 g(\mathbf{x})}{\partial x_2 \partial x_n} \\ ... & ... & ... & ... \\ \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_2} & ... & \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix} = 2A$$

(d) Geometry

i. Let$\mathbf{x}_1$,$\mathbf{x}_2$be two different points on the line $\mathbf{w}^T\mathbf{x} + b = 0$.
Then $\mathbf{x}_2 - \mathbf{x}_1$is a vector with the same direction of the line.
$\because \mathbf{w}^T\left(\mathbf{x}_2 - \mathbf{x}_1\right) = \mathbf{w}^T\mathbf{x}_2 - \mathbf{w}^T\mathbf{x}_1 = (-b) - (-b) = 0$
$\therefore$ $\mathbf{w}$ is orthogonal to $\mathbf{x}_2 - \mathbf{x}_1$
$\therefore$ $\mathbf{w}$ is orthogonal to the line $\mathbf{w}^T\mathbf{x} + b = 0$

ii. Let $\mathbf{x}_p$be a point on the line such that the vector from origin to $\mathbf{x}_p$is orthogonal to the line.
$\therefore\mathbf{x}_p = \lambda\mathbf{w}$,$\lambda \neq 0$
$\because$ $\mathbf{x}_p$is a point on the line
$\therefore$ $\mathbf{w}^T\mathbf{x}_p + b = 0$
$\therefore\mathbf{w}^T\lambda\mathbf{w} + b = 0 \Rightarrow \lambda\mathbf{w}^T\mathbf{w} + b = 0 \Rightarrow \lambda\|\mathbf{w}\|_2^2 + b = 0 \Rightarrow \lambda = \frac{-b}{||\mathbf{w}||_2^2}$
$\therefore$The distance $= |\mathbf{x}_p| = |\lambda|||\mathbf{w}||_2 = \frac{|b|}{||\mathbf{w}||_2}$

10. Sampling from a Distribution

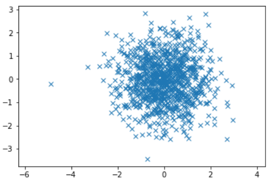
(a)

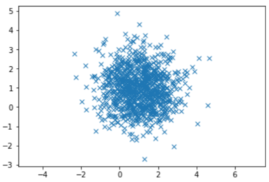
(b)

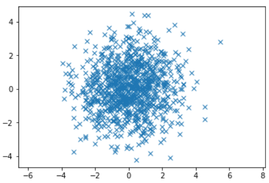
(c)

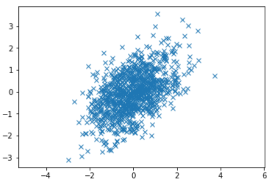
(d)

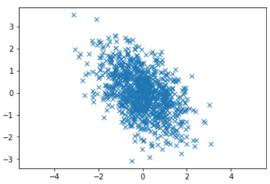
(e)


3
2
1
0
−1
−2
−3
−6 −4 −2 0 2 4


5
4
3
2
1
0
−1
−2
−3
−4 −2 0 2 4 6


4
2
0
−2
−4
−6 −4 −2 0 2 4 6 8


3
2
1
0
−1
−2
−3
−4 −2 0 2 4 6


3
2
1
0
−1
−2
−3
−4 −2 0 2 4