Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2019

Problem Set 0

Issued: Thursday 12$^{\text{th}}$ September, 2019 **Due:** Monday 16$^{\text{th}}$ September, 2019

Tips: It is not a formal homework and will not be graded. The primary goal is to help you remember those basic mathematics you have learnt before.

Probability Theory Part

0.1. (Conditional Probability) For discrete random variables, the conditional probability can be derived by Product Rule.

$$p(X, Y) = p(Y|X)\, p(X)$$

We can define the conditional expectation as

$$\mathbb{E}[Y|X = x] \triangleq \sum_{y \in \mathcal{Y}} y \cdot p(Y = y|X = x)$$

Explain that

(a) $\mathbb{E}[X|X] = X$

(b) $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}(X)$

(c) $\mathbb{E}[g(X)h(Y)|Y] = h(Y)\mathbb{E}[g(X)|Y]$ $\quad g(x)$ and $h(y)$ are bounded functions

0.2. (Bayes) A city has a 50% chance to rain everyday and the weather report has a 90% chance to correctly forecast.
You will take an umbrella when the report says it will rain and you have a 50% chance to take an umbrella when the report says it will not rain.
Compute

(a) the probability of raining when you don't take an umbralla

(b) the probability of not rainning when you take an umbrella

0.3. (Joint Distribution) Random Variables X and Y have joint distribution with joint probability density function

$$f(x, y) = \begin{cases} Ce^{-(2x+y)} & x > 0, y > 0 \\ 0 & ow. \end{cases}$$

Please find C by

$$\int_0^\infty \int_0^\infty f(x, y) \mathrm{d}x \mathrm{d}y = 1$$

0.4. (Covariance) For two random variables X and Y, the covariance is defined by

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Now we have a joint pdf

$$f(x, y) = \begin{cases} 4xy & 0 < x < 1, 0 < y < 1 \\ 0 & ow. \end{cases}$$

Please show that the covariance of X and Y is 0.

0.5. (Uncorrelated and independent RVs) We have a uniform distribution of X and Y on a disk. The pdf is

$$f(x, y) = \frac{1}{\pi} \qquad x^2 + y^2 \leq 1$$

When the covariance of X and Y is 0, we call them uncorrelated variables.
For continuous random variables, when the joint pdf can be written as the product of two RVs' pdf

$$f(x, y) = f_X(x) f_Y(y),$$

we call them independent.
Please show that X and Y are uncorrelated but not independent

0.6. (Guassian Distribution) There is a famous integral here

$$\int_{-\infty}^{\infty} e^{-x^2} \mathrm{d}x = \sqrt{\pi}$$

It is called Guassian Integral. Based on it, please find some results of the Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad -\infty < x < \infty$$

(a) Prove it is a pdf($\sigma > 0$)

(b) Compute the expectation and variance

Just hopefully you know Guassian Integral and remember these results. They are useful.

Calculus & Linear Algebra

0.7. (Chain rule) $x \in \mathbb{R}$ is a scalar, we have

$$\begin{aligned} y &= ax + b \\ z &= \frac{1}{1 + e^{-y}} \end{aligned} \tag{1}$$

Please give the $\frac{\partial z}{\partial x}$.

0.8. (Orthogonal) The $Q \in \mathbb{R}^{n \times n}$ is said to be **orthogonal** if its columns are pairwise orthogonal, which implies that

$$QQ^\top = Q^\top Q = I \tag{2}$$

Please show that $\|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$.

0.9. (Inner product) If $\boldsymbol{x} \in \mathbb{R}^n$ is orthogonal to $\boldsymbol{y} \in \mathbb{R}^n$, please show that

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \tag{3}$$

0.10. (Determinant) If a matrix $A \in \mathbb{R}^{n \times n}$ is invertible, the A^* is said to be the **adjoint matrix** of A where $A^{-1} = \frac{A^*}{\det A}$. Please prove that if $\det A = 0$, then we have $\det A^* = 0$.

0.11. (Invertibility) Given a matrix $A \in \mathbb{R}^{n\times n}$ and $A^3 = 4I$, please give the invertible matrix of $A - I$.

0.12. (Trace) The **trace** of a matrix $A \in \mathbb{R}^{n\times n}$ is defined as sum of diagonal elements of A:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii} \tag{4}$$

(a) Show that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

(b) Show that $\nabla_A \mathrm{tr}(AB) = B^\top$.

0.13. (Eigenthings) Let $\boldsymbol{x}$ be an eigenvector of a matrix A with corresponding eigenvalue λ, then

(a) Show that for any $\gamma \in \mathbb{R}$, the $\boldsymbol{x}$ is an eigenvector of $A+\gamma I$ with eigenvalue $\lambda+\gamma$.

(b) If A is invertible, then $\boldsymbol{x}$ is an eigenvector of A^{-1} with eigenvalue λ^{-1}.

(c) $A^k\boldsymbol{x} = \lambda^k\boldsymbol{x}$ for any $k \in \mathbb{Z}$ ($A^0 = I$ by definition)

0.14. (Matrix derivative) $\boldsymbol{x}, \boldsymbol{w} \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n\times n}$. We have $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{w}^\top \boldsymbol{x} \tag{5}$$

Please give the $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$.