

# Probability Theory Review

Student name: *Fabienne Guertler 12-935-508*

Course: *Machine Learning*  
Due date: *October 9th, 2019*

## Question 1 (15 points)

Show that the covariance matrix is always symmetric and positive semidefinite.

**Solution:** A matrix  $A$  is positive semidefinite if  $x^T A x \geq 0$  for all  $x \in \mathbb{R}^n$ .  
Covariance matrix  $\Sigma = \text{cov}[X, X] = E[XX^T] + E[X]E[X]^T$  with  $X \in \mathbb{R}^n$ .

$$\begin{aligned} u^T \Sigma u &= u^T (E[XX^T] + E[X]E[X]^T) u \\ &= u^T E[XX^T] u + u^T E[X]E[X]^T u \end{aligned}$$

## Question 2 (20 points)

$X \in \mathbb{R}^n$  and  $Y \in \mathbb{R}^m$  are independent random variables. Their expectations and covariances are  $E[X] = 0$ ,  $\text{Cov}[X] = I$ ,  $E[Y] = \mu$ , and  $\text{Cov}[Y] = \sigma I$ , where  $I$  is the identity matrix of the appropriate size and  $\sigma$  is a scalar. What is the expectation and covariance of the random variable  $Z = AX + Y$ , where  $A \in \mathbb{R}^{m \times n}$ ?

## Question 3 (15 points)

Thomas and Viktor are friends. It is Friday night and Thomas does not have a phone. Viktor knows that there is a  $2/3$  probability that Thomas goes to the party to downtown. There are 5 pubs in downtown and there is an equal probability of Thomas going to any of them if he goes to the party. Viktor already looked for Thomas in 4 of the bars. What is the probability of Viktor finding Thomas in the last bar?

## Question 4 (20 points)

Derive the mean for the Beta Distribution, which is defined as

$$\text{Beta}(x|a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}$$

where  $B(a, b)$ ,  $\Gamma(a)$  are Beta and Gamma functions respectively:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

**Solution:** To get integral representation of the beta function

$$\begin{aligned}
 B(a, b)\Gamma(a + b) &= \Gamma(a)\Gamma(b) \\
 &= \int_0^\infty e^{-u} u^{a-1} du \cdot \int_0^\infty e^{-v} v^{b-1} dv \\
 &= \int_{v=0}^\infty \int_{u=0}^\infty e^{-u-v} u^{a-1} v^{b-1} du dv \\
 u = f(z, t) = zt, \quad v = g(z, t) = z(1 - t) \\
 &= \int_{z=0}^\infty \int_{t=0}^\infty e^{-z} (zt)^{a-1} (z(1-t))^{b-1} |J(z, t)| dt dz \\
 &= \int_{z=0}^\infty \int_{t=0}^\infty e^{-z} (zt)^{a-1} (z(1-t))^{b-1} z dt dz \\
 &= \int_{z=0}^\infty e^{-z} z^{a+b-1} dz \cdot \int_{t=0}^1 t^{a-1} (1-t)^{b-1} dt
 \end{aligned}$$

We know that  $\Gamma(a + b) = \int_0^\infty e^{-z} z^{a+b-1} dz$  so we get

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$$

The mean of beta distribution

$$\begin{aligned}
 E(x) &= \int_0^1 x \text{Beta}(x|a, b) dx \\
 &= \int_0^1 x \frac{x^{a-1} (1-x)^{b-1}}{B(a, b)} dx \\
 &= \frac{1}{B(a, b)} \int_0^1 x^a (1-x)^{b-1} dx \\
 &= \frac{B(a+1, b)}{B(a, b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a\Gamma(a)\Gamma(b)\Gamma(a+b)}{(a+b)\Gamma(a+b)\Gamma(a)\Gamma(b)} = \frac{a}{a+b}
 \end{aligned}$$

### Question 5 (20 points)

Let  $A \in \mathbb{R}^{n \times n}$  be a positive definite square matrix,  $b \in \mathbb{R}^n$ , and  $c$  be a scalar. Prove that

$$\int_{x \in \mathbb{R}^n} e^{-\frac{1}{2}x^T A x - x^T b - c} dx = \frac{(2\pi)^{n/2} |A|^{-1/2}}{e^{c - \frac{1}{2}b^T A^{-1}b}}$$

### Question 6 (10 points)

From the definition of conditional probability of multiple random variables, show that

$$f(x_1, x_2, \dots, x_n) = f(x_1) \prod_{i=2}^n f(x_i | x_1, \dots, x_{i-1})$$

where  $x_1, x_2, \dots, x_n$  are random variables and  $f$  is a probability density function of its arguments.

**Solution:** The conditional probability density of multiple random variables

$$f_{X_n | X_1, \dots, X_{n-1}}(x_n | x_1, \dots, x_{n-1}) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{f_{X_1, \dots, X_{n-1}}(x_1, x_2, \dots, x_{n-1})}$$

Solve the equation for  $f(x_1, \dots, x_n)$

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f(x_n | x_1, \dots, x_{n-1}) f(x_1, \dots, x_{n-1}) \\ &= f(x_n | x_1, \dots, x_{n-1}) f(x_n | x_1, \dots, x_{n-2}) f(x_1, \dots, x_{n-2}) \\ &= f(x_n | x_1, \dots, x_{n-1}) f(x_n | x_1, \dots, x_{n-2}) f(x_1, \dots, x_{n-2}) \cdot \dots \cdot f(x_2 | x_1) f(x_1) \\ &= f(x_1) \prod_{i=2}^n f(x_i | x_1, \dots, x_{i-1}) \end{aligned}$$