UNIVERSITY OF BERN

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 $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$.
Covariance matrix $\Sigma = cov[X, X] = E[XX^T] + E[X]E[X]^T$ with $X \in \mathbb{R}^n$. Using $Cov[X, X] = Var[X]$ we get

$$u^T Var[X]u = Var[uX]$$

Because variance is always non-negative, we get $u^T Var[X]u \geq 0$. A matrix that is positive-semidefinite is automatically symmetric.

Question 2 (20 points)

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

Solution: Expectation for random variable Z is

$$E[Z] = E[AX + Y] = AE[X] + E[Y] = A \cdot 0 + \mu = \mu$$

the variance of Z

$$Var[Z] = Var[AX + Y] = Var[AX] + Var[Y] + 2Cov[X, Y]$$

Because X, Y are independent random variables $Cov[X, Y]$ is 0. This leads to

$$Var[Z] = Var[AX] + Var[Y] = AVar[X]A^T + Var[Y] = AIA^T + \sigma I = I(A^T A + \sigma)$$

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

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

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

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

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

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

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

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

Solution: We know that the integral of the Gaussian probability density function on a random variable with mean μ and covariance Σ is 1.

$$\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{x\in\mathbb{R}^n} \exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)) = 1$$

$$-\frac{1}{2}x^TAx - x^Tb - c = -\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b) - c + \frac{1}{2}b^TA^{-1}b$$

We can factor out the terms not including x

$$\begin{aligned}\int \exp(-\frac{1}{2}x^TAx - x^Tb - c)dx &= \int \exp(-\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b) - c + \frac{1}{2}b^TA^{-1}b)dx \\ &= \exp(-c + \frac{1}{2}b^TA^{-1}b) \cdot \int \exp(-\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b))dx\end{aligned}$$

We use $\int \exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx = (2\pi)^{n/2}|\Sigma|^{1/2}$ and $\Sigma = A^{-1}$ to get rid of the remaining integral and get

$$\int \exp(-\frac{1}{2}x^TAx - x^Tb - c)dx = \frac{(2\pi)^{n/2}|A|^{-1/2}}{e^{c-\frac{1}{2}b^TA^{-1}b}}$$

Question 6 (10 points)

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

$$f(x_1, x_2, ..., x_n) = f(x_1)\prod_{i=2}^{n} f(x_i|x_1, ..., x_n)$$

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

Solution: The conditional probability density of multiple random variables

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

Solve the equation for $f(x_1, ...x_n)$

$$\begin{aligned}f(x_1, x_2, ..., x_n) &= f(x_n|x_1, ..., x_{n-1})f(x_1, ..., x_{n-1}) \\ &= f(x_n|x_1, ..., x_{n-1})f(x_n|x_1, ..., x_{n-2})f(x_1, ..., x_{n-2}) \\ &= f(x_n|x_1, ..., x_{n-1})f(x_n|x_1, ..., x_{n-2})f(x_1, ..., x_{n-2}) \cdot ... \cdot f(x2|x_1)f(x1) \\ &= f(x_1)\prod_{i=2}^{n} f(xi|x1, ..., x_{i-1})\end{aligned}$$