Machine Learning Assignment # 3
Universität Bern

Due date: 09/10/2019

Late submissions will incur a penalty. Submit your answers in ILIAS (as a pdf or as a picture of your written notes if the handwriting is very clear). Submission instructions will be provided via email. You are not allowed to work with others.

Probability theory review **[Total 100 points]**

Solve each of the following problems and show all the steps of your working.

1. Show that the covariance matrix is always symmetric and positive semidefinite. **[15 points]**
   Solution
   The $(i,j)^{th}$ element of the covariance matrix Σ is given by

   $$\Sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E[(X_j - \mu_j)(X_i - \mu_i)] = \Sigma_{ji}$$

   so that the covariance matrix is symmetric.

   For an arbitrary vector u,

   $$\begin{aligned} u^T \Sigma u &= u^T E[(X-\mu)(X-\mu)^T]u = E[u^T (X-\mu)(X-\mu)^T u] \\ &= E[((X-\mu)^T u)^T (X-\mu)^T u] = E[((X-\mu)^T u)^2] \geq 0 \end{aligned}$$

   so that the covariance matrix is positive semidefinite.

2. $X \in R^n$ 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 $A \in R^{m \times n}$? **[20 points]**
   Solution.
   The expectation of Z can be obtained from the definition by applying the linearity of expectation,

   $$E[Z] = E[AX + Y] = AE[X] + E[Y] = 0 + \mu = \mu. \tag{1}$$

   The covariance of Z is $Cov[Z] = E[ZZ^T] - E[Z]E[Z] = E[ZZ^T] - \mu\mu^T$. Substituting the definition of Z, we get the expression below.

   $$\begin{aligned} E[ZZ^T] &= E[(AX+Y)(AX+Y)^T] = && (2) \\ &= E[AXX^TA^T + YX^TA^T + AXY^T + YY^T] = && (3) \\ &= AE[XX^T]A^T + E[YX^T]A^T + && (4) \\ &\quad AE[XY^T] + E[YY^T]. && (5) \end{aligned}$$

   Here we can substitute $E[XX^T] = I$ and $E[YY^T] = \sigma I + \mu\mu^T$. Because X and Y are independent, $E[XY^T] = E[X]E[Y]^T = 0$, similarly $E[YX^T] = 0$. We get $E[ZZ^T] = AA^T + \sigma I + \mu\mu^T$, therefore $Cov[Z] = AA^T + \sigma I$.

3. 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? **[15 points]**
Solution.
The sample space is

$$S = \{\text{"home"}, \text{"pub 1"}, \text{"pub 2"}, \text{"pub 3"}, \text{"pub 4"}, \text{"pub 5"}\}, \tag{6}$$

and the probability of the events are $P(\text{"home"}) = 1/3$ and $P(\text{"pub i"}) = 2/15$. We need to compute $P(\text{"pub 5"}|\text{"not in pub 1 ... 4"})$. Using the Bayes rule,

$$\begin{aligned} P(\text{"pub 5"}|\text{"not in pub 1 ... 4"}) &= \\ \frac{P(\text{"pub 5"} \cap \text{"not in pub 1 ... 4"})}{P(\text{"not in pub 1 ... 4"})} &= \frac{2/15}{7/15} = \frac{2}{7}. \end{aligned} \tag{7, 8}$$

4. Derive the mean for the Beta Distribution, which is defined as **[20 points]**

$$\text{Beta}(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \tag{9}$$

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

$$B(a,b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{10}$$

and

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1}e^{-u}du. \tag{11}$$

Hint: *Use integration by parts.*
Solution.

$$E[X] = \int_0^1 x \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} dx \tag{12}$$

To solve this problem, we will integrate by parts.

$$\begin{aligned} E[X] &= \frac{1}{B(a,b)} \left[\frac{x^a(-1)(1-x)^b}{b} \bigg|_0^1 - \int_0^1 a x^{a-1}(-1)\frac{(1-x)^b}{b} dx \right] = \\ &= \frac{1}{B(a,b)} \int_0^1 \frac{a}{b} x^{a-1}(1-x)^{b-1}(1-x)dx \end{aligned} \tag{13}$$

$$\begin{aligned} E[X] &= \frac{1}{B(a,b)} \int_0^1 \frac{a}{b} x^{a-1}(1-x)^{b-1}(1-x)dx = \\ &= \frac{1}{B(a,b)} \left[\int_0^1 \frac{a}{b} x^{a-1}(1-x)^{b-1}dx - \int_0^1 x\frac{a}{b} x^{a-1}(1-x)^{b-1}dx \right] = \\ &= \frac{a}{b}(1 - E[X]) \end{aligned} \tag{14}$$

$$E[X] = \frac{a}{a+b} \tag{15}$$

5. 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 **[20 points]**

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

Hint: *Use the fact that the integral of the Gaussian probability density function of a random variable with mean μ and covariance Σ is 1.*
Solution

$$\begin{aligned} &= \int_{x\in\mathbb{R}^n} \exp(-\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b) + \frac{1}{2}b^TA^{-1}b - c)dx \\ &= \int_{x\in\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}|A^{-1}|^{1/2}} \exp(-\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b) + \frac{1}{2}b^TA^{-1}b - c)dx * (2\pi)^{n/2}|A^{-1}|^{1/2} \\ &= \int_{x\in\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}|A^{-1}|^{1/2}} \exp(-\frac{1}{2}(x + A^{-1}b)^TA(x + A^{-1}b))dx * \exp(\frac{1}{2}b^TA^{-1}b - c) * (2\pi)^{n/2}|A^{-1}|^{1/2} \\ &= \frac{(2\pi)^{n/2}|A|^{-1/2}}{\exp(c - \frac{1}{2}b^TA^{-1}b)} \end{aligned}$$

6. From the definition of conditional probability of multiple random variables, show that **[10 points]**

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

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

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