Review Assignment 1

Nalet Meinen
Machine Learning

October 9, 2019

1   Probability theory review

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.

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,

$$u^\mathsf{T} \Sigma u = u^\mathsf{T} E[(X - \mu)(X - \mu)^\mathsf{T}]u = E[(u^\mathsf{T}(X - \mu)(X - \mu)^\mathsf{T})u]$$
$$= E[((X - \mu)^\mathsf{T}u)^\mathsf{T}(X - \mu)^\mathsf{T}u] = E[((X - \mu)^\mathsf{T}u)^2] \geqslant 0$$

so that the covariance matrix is positive semi-definite.

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

The expectation of Z can be obtained from the definition by applying the linearity of expectation,

$$E[Z] = E[AX + Y] = AE[x] + E[Y] = 0 + \mu = \mu$$

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

$$E[ZZ^{\mathsf{T}}] = E[(AX + Y)(AX + Y)^{\mathsf{T}}] =$$
$$= E[AXX^{\mathsf{T}}A^{\mathsf{T}} + YX^{\mathsf{T}}A^{\mathsf{T}} + AXY^{\mathsf{T}} + YY^{\mathsf{T}}] =$$
$$= AE[XX^{\mathsf{T}}]A^{\mathsf{T}} + E[YX^{\mathsf{T}}]A^{\mathsf{T}} + AE[XY^{\mathsf{T}}] + E[YY^{\mathsf{T}}]$$

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

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 ending Thomas in the last bar?

The sample space is

$$S = f\{\text{home}, \text{pub } 1, \text{pub } 2, \text{pub } 3, \text{pub } 4, \text{pub } 5\}$$

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

$$P(\text{pub } 5|\text{not in pub } 1 \ldots 4) =$$
$$\frac{P(\text{pub } 5 \cap \text{not in pub } 1 \ldots 4)}{P(\text{not in pub } 1 \ldots 4)} = \frac{\frac{2}{15}}{\frac{7}{15}} = \frac{2}{7}$$

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

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

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

$$B(a, b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} \tag{2}$$

and

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1}e^{-u}du \tag{3}$$

Hint: Use integration by parts.

$$E(x) = \frac{1}{\text{Beta}(a,b)} \int_0^1 x^a (1-x)^{b-1} dx$$

$$= \frac{\text{Beta}(a+1,b)}{\text{Beta}(a,b)}$$

$$= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

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

$$= \frac{a}{a+b}$$

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

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

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.

We know from theory: $f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^\mathsf{T}|\Sigma|^{-1}(x-\mu)}$

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

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

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

$$= (2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}} \cdot e^{\frac{1}{2}b^\mathsf{T} A^{-1}b - c} \cdot 1$$

$$= \frac{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}{e^{\frac{1}{2}b^\mathsf{T} A^{-1}b - c}}$$

We can replace Σ with the ones corresponding to the formula.

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

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

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

$$P(x_1 \cap x_2 \cap \cdots \cap x_n) = P(x_1) \prod_{i=2}^{n} P(x_i | x_1, \ldots x_{i-1})$$

When $n = 2$

$$P(x_1 \cap x_2) = P(x_1) \prod_{i=2}^{n} P(x_2 | x_1)$$
$$= P(x_1) P(x_2 | x_1)$$

$$P(x_1 \cap x_2 \cap \cdots \cap x_{n-1} \cap x_n) = P(x_n | x_1 \cap x_2 \cap \cdots \cap x_{n-1} \cap x_n) P(x_1 \cap x_2 \cap \cdots \cap x_{n-1} \cap x_n)$$