Machine Learning, Tutorial 1
University of Bern

Abdelhak Lemkhenter (abdelhak.lemkhenter@inf.unibe.ch)

18/09/2019

Linear algebra

1. Given $A \in \mathbb{R}^{n \times n}$, show that if $\forall M \in \mathbb{R}^{n \times n}, MA = AM$ then $\exists \lambda \in \mathbb{R}$ such that $A = \lambda Id$.
Hint: Use matrices $(B^i)_{i \in \{1,...,n\}}$, $B^i := diag(\lambda_1, ..., \lambda_n)$ where $\lambda_k = 1$ for $k \neq i$ and $\lambda_i = 2$.

Solution. Let's define $X = AB^1$ and $Y = B^1 A$. Multiplying by B^1 on the left and on the right is scaling by a factor of 2 the first column and row of A respectively.

$$\begin{aligned} X = Y &\Rightarrow X_{1,j} = Y_{1,j}, \forall j \in \{1, ..., n\} \\ &\Rightarrow 2A_{j,1} = A_{j,1} \wedge A_{1,j} = 2A_{1,j}, \forall j \in \{2, ..., n\} \\ &\Rightarrow A_{j,1} = 0 \wedge A_{1,j} = 0 \forall j \in \{2, ..., n\} \end{aligned}$$

Alternatively, multiplying by B^i on the left and on the right is scaling by a factor of 2 the i-th column and row of A respectively. Therefore, all non-diagonal terms pf A are zero. We can rewrite A as $diag(\lambda_1, ..., \lambda_n)$. Let's take an arbitrary matrix $M = [m_1 \dots m_n] = \begin{bmatrix} m_1^T \\ \dots \\ m_n^T \end{bmatrix}$. Then $MA = [\lambda_1 m_1 \dots \lambda_n m_n] = AM = \begin{bmatrix} \lambda_1 m_1^T \\ \dots \\ \lambda_n m_n^T \end{bmatrix}$ which is true only if all the diagonal elements of A are equal. Therefore $\exists \lambda \in \mathbb{R}$ such that $A = \lambda Id$.

2. Show that a positive definite matrix is non-singular.
Solution. Suppose A is positive definite and non-singular. So it is not full rank.Then there is a column of A, like j, that can be expressed as a linear combination of the other columns.

$$a_j = \sum_{i \neq j} x_i a_i$$

setting $x_j = -1$, we have

$$Ax = \sum_{i=1}^{n} x_i a_i = 0$$

But this implies that $x^\top Ax = 0$

3. Show that if (λ_i, x_i) are the i-th eigenvalue and i-th eigen vector of a non-singular and symmetric matrix $A \in \mathbb{R}^{n \times n}$, then $(\frac{1}{\lambda_i}, x_i)$ are the i-th eigenvalue and i-th eigen vector of A^{-1}.

Solution.

$$Ax = \lambda_i x_i \rightarrow x_i = \lambda_i A^{-1} x_i \rightarrow A^{-1} x_i = \frac{1}{\lambda_i} x$$

4. Show that if $U \in \mathbb{R}^{n \times n}$ is orthogonal than $U^{-1} = U^T$.

Solution.

$$(U^T U)_{i,j} = u_i^T u_j = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

5. Show that, for a symmetric matrix A, if x_1, x_2 are two eigen vectors of A such that their respective eigen values are distinct and non-zero then $x_1^T x_2 = 0$.

   Solution.

$$\begin{aligned} x_2^T A x_1 &= x_2^T \lambda_1 x_1 \\ &= \lambda_1 x_2^T x_1 \\ &= (x_2^T A x_1)^T \\ &= x_1^T A x_2 \\ &= \lambda_2 x_1^T x_2 \Rightarrow x_1^T x_2 = 0 \end{aligned}$$

6. Show that rank$(A) \leq \min\{m, n\}$, where $A \in \mathbb{R}^{m \times n}$.

   Solution. We know that column rank and row rank of any matrix is equal. also we know that the column rank is at most equal to the number of columns and the row rank is at most equal to the number of rows. These two consideration implies that rank$(A) \leq \min\{m, n\}$.

7. In each of the following cases, state whether the matrix A is guaranteed to be non-singular or not. Justify your answer in each case.

   (a) $A \in \mathbb{R}^{m \times n}$ is a full rank matrix.
   Solution. No. A non-singular matrices should be square.

   (b) $|A| = 0$.
   Solution. No. An square matrix A is non-singular if and only $|A| \neq 0$

   (c) A is an orthogonal matrix.
   Solution. Yes, For an orthogonal matrix Q we have $Q^\top Q = QQ^\top = I$ so Q is non singular and $Q^{-1} = Q^\top$

   (d) A has no eigenvalue equal to zero.
   Solution. Yes, we know that $|A| = \prod \lambda_i$. So if A has no zero eigenvalue, then $|A| \neq 0$ so A is non-singular.

   (e) A is a symmetric matrix with non-negative eigenvalues.
   Solution. No. we know that if A is a symmetric matrix then $x^\top A = \sum_{i=1}^{n} \lambda_i y_i^2$ and it is positive/negative for any x if and only if all the eigenvalues are positive/negative.

Probability

1. In this exercise we analyse the Monty Hall game. The rules of the game are the following. There are 3 doors A, B and C. Behind one of the doors there is the great prize. Behind other doors there is nothing. Firstly, the player chooses one of the doors. Secondly the anchorman (Monty Hall) opens one of the doors the player did not choose, such that the prize is not behind that door. The anchorman can always do this, because he knows where the prize is. At this point the player has to choose whether he sticks with the first choice, or pick the other door. The player will receive whatever behind the door she chooses. What is the best strategy to win the prize?

   Solution.
   Without loss of generality, we assume the player chose door #1 and Monty Hall opens door #2. We define S, G as the event where the prize is behind door #1 and the event Monty Hall opened a door with no prize. Using the Bayes formula:

$$\begin{aligned} P(A|G) &= \frac{P(G|A)P(A)}{P(G)} \\ &= \frac{P(G|A)P(A)}{P(G|A)P(A) + P(G|\sim A)P(\sim A)} \\ &= \frac{1 * 1/3}{1 * 1/3 + 1 * 2/3} = 1/3 \end{aligned}$$

   Therefore $P(\sim A|G) = 2/3$, so the optimal strategy is to change your choice.

2. Given the following statistics what is the probability that a woman has cancer if she has a positive mammogram result?

   - One percent of women over 50 have breast cancer.

- Ninety percent of women who have breast cancer test positive on mammograms.
- Eight percent of women will have false positives.

Solution. We define H as the event where the patient has cancer and e as the event where the patient tests positive.

$$\begin{aligned} P(H|e) &= \frac{P(H,e)}{P(e)} = \frac{P(e|H)P(H)}{P(e|H)P(H) + P(e|\sim H)P(\sim H)} \\ &= \frac{0.9 \times .01}{.9 \times .01 + .08 \times 0.99} \approx 0.1 \end{aligned}$$

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

Solution.
The sample space is

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

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

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

Note that :

$$P(\text{"not in pub 1 ... 4"}) = P(\text{"home"} and \text{"not in pub 1 ... 4"}) + P(\text{"out"} and \text{"not in pub 1 ... 4"}) = \frac{1}{3} \times 1 + \frac{2}{3} \times \frac{1}{5} = \frac{7}{15}$$

4. Prove the following statements regarding the co-variance:

(a) Show that if X and Y are independent then $\text{Cov}[X,Y] = 0$.
Give an example that shows that the opposite is not true.
Solution

$$\begin{aligned} \text{Cov}[X,Y] &= E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y] \end{aligned}$$

If X, Y are independent, $E[XY] = E[X]E[Y]$ therefore the co-variance is equal to 0. To show that the opposite is not correct, consider random variables $X \sim \mathcal{N}(0,1), Y = X^2$ we know that $E(X) = E(X^3) = 0$. Therefore, $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X^3] - E[X]E[X^2] = 0$. However, it is clear that X, Y are not independent.

(b) Show that the covariance matrix is always symmetric and positive semi-definite.
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,

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

so that the covariance matrix is positive semi-definite.

5. $X \in R^n$ and $Y \in R^m$ are independent random vectors. Their expectations and covariance matrices 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 scalar. What is the expectation and covariance matrix of the random vector $Z = AX + Y$, where $A \in R^{m \times n}$?

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

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

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

$$\begin{aligned} E[ZZ^\top] &= E[(AX + Y)(AX + Y)^\top] = \\ &= E[AXX^\top A^\top + YX^\top A^\top + AXY^\top + YY^\top] = \\ &= AE[XX^\top]A^\top + E[YX^\top]A^\top + AE[XY^\top] + E[YY^\top]. \end{aligned}$$

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