Exercise Set (due on Sunday 17th Nov 2019)
Elements of Mathematics – Bioinformatics for Health Sciences

1. Let

$$M = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}.$$

(a) Is M invertible? Justify your answer.

Answer: Yes, it is. This is a square matrix with maximum rank. You can also check that it has non-zero determinant: $\det(M) = -\frac{1}{5} - \frac{4}{5} = -1 \neq 0$.

(b) Compute M^{-1} using Gauss-Jordan elimination.

Answer: First notice that $M = \frac{1}{\sqrt{5}} M_0$ where

$$M_0 = \begin{bmatrix} 1 & 2 \\ 2 & \text{-1} \end{bmatrix}$$

Then $M^{-1} = \sqrt{5} M_0^{-1}$. Let's compute M_0^{-1} (for convenience I will highlight the right hand side matrix in yellow):

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{bmatrix} \overset{i}{\sim} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -5 & \text{-2} & 1 \end{bmatrix} \overset{ii}{\sim} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 2/5 & \text{-1/5} \end{bmatrix} \overset{iii}{\sim} \begin{bmatrix} 1 & 0 & 1/5 & 2/5 \\ 0 & 1 & 2/5 & \text{-1/5} \end{bmatrix},$$

with steps:

i. $R_2 \leftarrow R_2 - 2 \cdot R_1$,
ii. $R_2 \leftarrow (-1/5) \cdot R_2$,
iii. $R_1 \leftarrow R_1 - 2 \cdot R_2$.

Then

$$M^{-1} = \sqrt{5} \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & -1/5 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} = M.$$

(c) Compute $\det(M)$ and $\det(M^{-1})$.

Answer: Since $M = M^{-1}$, $\det(M) = \det(M^{-1}) = -1$.

2. Notice that the determinant turns matrix multiplication into ordinary product of real numbers, that is, if $A, B \in M_{n\times n}$ are square matrices, then $\det(AB) = \det(A)\det(B)$.

Consider the matrices:

$$A = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 2 & 2 \\ 0 & 0 & 7 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & -2 \end{bmatrix}$$

(a) Using the recursive definition of determinant, reason why the determinant of a diagonal matrix is the product of the entries in the diagonal, i.e. if $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ then $\det(D) = \lambda_1 \cdot \ldots \cdot \lambda_n$.

Answer: The matrix $D = (d_{ij})$ has entries $d_{ii} = \lambda_i$ and $d_{ij} = 0$ for $i \neq j$. The recursive definition of determinant we saw in class reads as follows:

$$\det D = \sum_{i=1}^{n} (-1)^{i+1} d_{i1} \det(D_{-i1}),$$

where D_{-i1} is the submatrix of D upon removing the i-th row and 1-st column. Since $d_{i1} = 0$ for all $i \neq 1$ and $\det(D_{-11}) = \text{diag}(\lambda_2, \dots, \lambda_n)$, the previous recurrence simplifies considerably:

$$\det D = \lambda_1 \cdot \det(\text{diag}(\lambda_2, \dots, \lambda_n)).$$

Clearly the same argument used to compute the determinant of the smaller diagonal matrix, until we reach the base case where we have to compute the determinant of the 1×1 matrix $[\lambda_n]$.

(b) Compute the determinant of the identity matrix.
Answer: Accoding to the previous remark $\det(\text{Id}_n) = 1^n = 1$.

(c) Notice that the matrices A and B given above satisfy that all their entries below the diagonal are zero: matrices satisfying this condition are known as "upper triangular" matrices. Reason why the determinant of an upper triangular matrix is the product of the entries in the diagonal, i.e. if $T = (a_{ij})$ is upper triangular, then $\det(T) = a_{11} \cdot \ldots \cdot a_{nn}$.
Answer: By definition

$$\det T = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(T_{-i1}),$$

where $a_{i1} = 0$ for all $i \neq 1$ and T_{-11} is upper-triangular, then

$$\det T = a_{11} \cdot \det(T_{-11}).$$

Now the same line of reasoning can be as in part i) can be reproduced.

(d) Verify that the opening remark holds for the matrices A and B given above, that is, $\det(AB) = \det(A)\det(B)$.
Answer: You can compute the three determinants and check the relation holds. Another avenue is to reason why this is true in the case of upper-triangular matrices. First observe then that the product of upper-triangular matrices has to yield an upper-triangular matrix. By the previous part, to compute the determinant of AB we are only concerned about its diagonal entries: its i-th diagonal entry is given by

$$(AB)_{ii} = \sum_{k=1}^{n} a_{ik} b_{ki} = a_{ii} b_{ii}$$

since $a_{ik} = 0$ for $k < i$ and $b_{ki} = 0$ for $k > i$. Consequently, $\det(AB) = (a_{11}b_{11}) \cdot \ldots \cdot (a_{nn}b_{nn})$. Then it is clear that

$$\det(AB) = (a_{11} \cdot \ldots \cdot a_{nn})(b_{11} \cdot \ldots \cdot b_{nn}) = \det(A)\det(B).$$

(e) Recall that when A is an invertible matrix, it satisfies $AA^{-1} = A^{-1}A = I$. Making use of the opening remark if necessary, reason why $\det(A^{-1}) = 1/\det(A)$.
Answer: Since $AA^{-1} = A^{-1}A = I$ by applying determinant we have $\det(AA^{-1}) = \det(A)\det(A^{-1}) = \det(\text{Id}_m) = 1$. Making use of the last two equations we conclude that $\det(A^{-1}) = 1/\det(A)$. Also there is no problem in dividing by $\det(A)$ because any invertible matrix has non-zero determinant.

3. Let

$$A = \begin{bmatrix} -4 & 2 & 0 \\ 2 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Let $f_A : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map defined as $f_A(v) = Av$.

(a) What is the rank of the matrix A? Justify your answer.
Answer: $\text{rank}(A) \le 2$ because the first two rows are multiples of each other, but since the last two rows are linearly independent (why?) $\text{rank}(A) \ge 2$. Then $\text{rank}(A) = 2$.

(b) Let $\{e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)\}$ be the canonical basis of $\mathbb{R}^3$.
Compute $f(e_1)$, $f(e_2)$ and $f(e_3)$.
Answer: Observe that $f(e_1)$, $f(e_2)$ and $f(e_3)$ are just the column vectors of A.

(c) Give a basis of the vector subspace $S \subset \mathbb{R}^3$ generated by $f(e_1), f(e_2), f(e_3)$.
Answer: Since $\text{rank}(A) = 2$ the columns of A do not form a linearly independent set., but two of them – not necessarily any two of them – can. In this case $f(e_1)$, $f(e_2)$ generate S and are linearly independent, consequently they are a basis of S.

4. The transpose matrix of $A = (a_{ij}) \in M_{n\times m}$ is another matrix that has as rows the columns of A: it is denoted A^t. In particular, notice that $A^t \in M_{m\times n}$. For example:

$$A = \begin{bmatrix} -1 & 1 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \quad A^t = \begin{bmatrix} -1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

Taking the matrix A in the example:

(a) Can you deduce the sizes of AA^t and A^tA without doing any computation?
Answer: AA^t has size 3×3, while A^tA has size 2×2.

(b) Compute AA^t and A^tA.
Answer:

$$AA^t = \begin{bmatrix} 2 & -3 & 0 \\ -3 & 5 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad AA^t = \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$$

(c) We say that a square matrix M is "symmetric" if $A^t = A$. Are AA^t and A^tA symmetric?
Answer: Yes, they are. The matrix AA^t represents all the possible products $r_i^t r_j$ where r_i are the rows of A. The matrix A^tA represents all the possible products $c_i^t c_j$ where c_i are the columns of A. The symmetry of these matrices is just a consequence of the fact that $v^t w = w^t v$ for any column vectors v, w of the same size (see exercise below).

5. Given two data vectors $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ of the same size, we define the mean $\text{E}(X)$, variance $\text{Var}(X)$ and covariance $\text{Cov}(X,Y)$ as:

$$\text{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))^2$$

$$\text{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

Informally, the **mean** embodies the idea of "center of mass" of the entries of a vector, the **variance** measures the extent to which a set of values tends to depart away from their mean; the **covariance** measures the extent to which two collections of values tend to depart from the respective means concordantly.

Given the following dummy data table:

	X	Y	Z
sample 1	-1	3	5
sample 2	0	6	3
sample 3	1	9	1

(a) Compute $\mathrm{E}(X)$, $\mathrm{E}(Y)$ and $\mathrm{E}(Z)$.

Answer: $\mathrm{E}(X) = 0$, $\mathrm{E}(Y) = 6$, $\mathrm{E}(Z) = 3$.

(b) Compute $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$ and $\mathrm{Var}(Z)$.

Answer: $\mathrm{Var}(X) = 1/3(1^2 + 1^2) = 2/3$, $\mathrm{Var}(Y) = 1/3(3^2 + 3^2) = 6$, $\mathrm{Var}(Z) = 1/3(2^2 + 2^2) = 8/3$.

(c) Compute $\mathrm{cov}(X, Y)$.

Answer: $\mathrm{cov}(X, Y) = 1/3\,[(-1)\cdot(-3) + 0\cdot 0 + 1\cdot 3] = 2$

(d) Let

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Compute $\mathbf{1}\mathbf{1}^t$.

Answer:

$$\mathbf{1}\mathbf{1}^t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(e) Let

$$Y = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

and define a new vector

$$C_Y = Y - \mathrm{E}(Y)\mathbf{1}.$$

Compute $\mathrm{E}(C_Y)$.

Answer: $\mathrm{E}(C_Y) = \mathrm{E}(Y - \mathrm{E}(Y)\mathbf{1}) = \mathrm{E}(Y) - \mathrm{E}(\mathrm{E}(Y)\mathbf{1}) = \mathrm{E}(Y) - \mathrm{E}(Y) = 0$.

(f) Let

$$C_3 = I_3 - \frac{1}{3}\mathbf{1}\mathbf{1}^t,$$

where I_3 is the 3×3 identity matrix. Compute C_3. Check that $C_3 Y = C_Y$. Can you guess why the matrix C_3 is known as the "centering matrix" of $\mathbb{R}^3$?

Answer:

$$C_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}.$$

It is apparent that the effect of multiplying a column vector by C_n is to center all its entries at the mean.

(g) Let

$$A = \begin{bmatrix} -1 & 3 \\ 0 & 6 \\ 1 & 9 \end{bmatrix}.$$

Observed that A encodes the two feature vectors corresponding to X and Y in the dummy table. Compute $B = C_3 A$. Explain in simple terms what is the effect that multiplying by C_3 has on A.

Answer:

$$B = C_3 A = \begin{bmatrix} -1 & -3 \\ 0 & 0 \\ 1 & 3 \end{bmatrix}.$$

(h) Compute

$$\omega = \frac{1}{3}B^tB.$$

ω is known as the "covariance" matrix. Discuss whether you see any connections between the entries of ω and the definitions of **variance** and **covariance** above.

Answer:

$$\omega = \frac{1}{3}B^tB = \frac{1}{3}\begin{bmatrix}2 & 6\\ 6 & 18\end{bmatrix} = \begin{bmatrix}2/3 & 2\\ 2 & 6\end{bmatrix}.$$

It turns out that $\omega_{11} = \mathrm{Var}(X)$, $\omega_{12} = \omega_{21} = \mathrm{Cov}(X, Y)$, and $\omega_{22} = \mathrm{Var}(Y)$.

6. The length of a vector

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$$

denoted $\|v\|$, can be defined as $\|v\| = \sqrt{v^tv}$

(a) Compute the lengths of $v = (1, 1, 0)$ and $w = (1, 2, 0)$, respectively.
Answer: $\|v\| = \sqrt{1^2 + 1^2} = \sqrt{2}$, $\|w\| = \sqrt{1^2 + 2^2} = \sqrt{5}$.

(b) Find a scalar $\alpha \in \mathbb{R}$ such that $\|\alpha v\| = 1$. Do the same for w.
Answer: Notice that in general $\|\alpha v\| = \sqrt{\alpha^2}\|v\| = |\alpha|\|v\|$. For v, $\alpha_v = 1/\|v\| = 1/\sqrt{2}$ is one possible solution. The same for w, $\alpha_w = 1/\|w\| = 1/\sqrt{5}$ is one possible solution. Are there any other solutions?

(c) Observe that given two vectors $v, w \in \mathbb{R}^n$ then $v^tw = w^tv$. Can you explain why?
Answer: Both expressions yield the following scalar:

$$\sum_{i=1}^{n} v_iw_i.$$

(d) We say that two vectors $v, w \in \mathbb{R}^n$ are orthogonal whenever $v^tw = 0$. Given

$$v = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \in \mathbb{R}^n$$

find two linearly independent vectors w_1 and w_2 such that both are orthogonal to v.
Answer: In this exercise we are bound to find vectors $w = (a, b, c)$ such that

$$a - 2c = 0.$$

We can provide any such vector with the constraint $a = 2c$. One simple solution is to make $a = c = 0$ and choose any $b \neq 0$, like $(0, 1, 0)$. Now, if we can given another non-zero solution with $b \neq 0$ we would have produced two linearly independent vectors that fulfill the orthogonality requirement. In fact, we could take $b = 0$ and $a \neq 0$, for instance, $a = 2$, then $c = 1$. In summary, we could just give $w_1 = (2, 0, 1)$ and $w_2 = (0, 1, 0)$.

7. **(Bonus Track)** Given the binary data $D = \{(x_i, y_i) \mid x_i, y_i \in \{0, 1\}, i = 1, \ldots, N\}$ we can fit a logistic regression model by maximizing the following log-likelihood function:

$$\mathcal{L}(a, b) = \sum_{i=1}^{N} -(ax_i + b) + (ax_i + b)y_i - \log(1 + e^{-(ax_i+b)})$$

where $a, b \in \mathbb{R}$ are the parameters of the model; in other words, solving the following optimization problem: $\hat{a}, \hat{b} = \mathrm{argmax}_{a,b}\mathcal{L}(a, b)$. In this exercise we are tackling this problem. For this purpose, recall that the binary data D can also be specified as a contingency table:

		y	
		0	1
x	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

where each $n_{ij} \in \mathbb{N}$ represents the count of data points that meet each of the four possible configurations for (x_i, y_i) in D: $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$.

(a) Prove the following identity:

$$\frac{\partial \mathcal{L}}{\partial a} = \sum_{i=1}^{N} x_i \left[y_i - \frac{e^{ax_i+b}}{1+e^{ax_i+b}} \right].$$

(b) Prove the following identity:

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{N} \left[y_i - \frac{e^{ax_i+b}}{1+e^{ax_i+b}} \right].$$

(c) Write $\partial \mathcal{L} / \partial a$ as a function of a, b and the n_{ij}.

(d) Write $\partial \mathcal{L} / \partial b$ as a function of a, b and the n_{ij}.

(e) Give an expression for the only critical point $(\hat{a}, \hat{b})$ of $\mathcal{L}$ as a function of the n_{ij}.