Elements of Computational Biology

Saul Pierotti

October 28, 2019

Introduction and topics

- Linear algebra is one of the main topics for developping algorithms
- Biology ha an high level of noise in its data
- I should know the formula of binomial and normal distribution
- We will study tests like T-Student, ANOVA
- Slides: ww.biocomp.unibo.it/gigi/2019-2020/ECB

Vectors

- Given a reference system, a vector is represented by its components on the axes
- The span of n vectors is the set of all possible vectors that you can represent by their linear combinations
- If a vector can be expressed as a linear combination of another, it is said to be linear dependent from it
- The basis of a vector space is a set of linearly independent vectors that span the full space
- $\vec{x} \in R^n$ means x is a real vector in an n-dimensional space
- $\vec{x} \in C^n$ means x is a complex vector in an n-dimensional space
- Sum of vectors is done by summing their components or graphically with the parallelogram rule
    - $\vec{c} = \vec{a} + \vec{b} \implies c_i = a_i + b_i \ \forall \ i = 1 \to n$
    - Difference is the same concept
    - $\forall \ \vec{v} \ \exists \ \vec{0} \ : \ \vec{v} + \vec{0} = \vec{v}$
    - You can only sum vectors in the same vector space
- The norm of a vector $||\vec{v}||$ is its lenght
    - Can be computed with the pitagorean theorem $||\vec{v}|| = \sqrt{\sum_{i=1}^{n} v_i^2}$
    - The norm of the sum is less or equal to the sum of the norm of the components
        * This follows from the geometry of a triangle
    - The scalar product of a norm is the norm of the scalar product
        * $\lambda ||\vec{v}|| = ||\lambda \vec{v}||$
- The distance between points in space is the norm of the difference between the vectors defining the points
    - $d(a, b) = ||\vec{a} - \vec{b}||$
- Scalar multiplication
    - $\vec{c} = \lambda \vec{a} \implies c_i = a_i \lambda$
    - A scalar multiplication of a sum is the sum of the scalar multiplications of the components
- Dot product, also called scalar or inner product
    - You can use the notation $< A, B >$
    - It is used in physics to calculate work
    - $\vec{w} = ||\vec{F}|| * ||\vec{s}|| * cos\theta = \sum_{i=1}^{n} F_i * s_i$
    - It is a number, complex or real depending on the vectors (!)
    - It is commutative and distributive
    - $< x, x > = ||\vec{x}||^2$
    - It is positive when the angle is acute

- – No cancelation rule
    - ∗ $< A, B > = < A, C > \not\Longrightarrow \vec{B} = \vec{C}$
- Angle between vectors
    - – Can be calculated inverting the dot product
- A line passing through the origin can be defined as the set of points orthogonal to a vector $\vec{w}$
    - – $w_1 x_1 + w_2 x_2 = 0$
    - – In higher dimensions this describes an hyperplane (an n-1 dimensional object)
- All the point on an hyperplane have the same projection on its defining vector $\vec{w}$
    - – The projection p of $\vec{x}$ on $\vec{w}$ is calculated as $\vec{x} * cos\theta$
    - – An hyperplane is therefore an object subjected to the constraint $\vec{x} * cos\theta = p$
    - – Given that $< \vec{x}, \vec{w} > = ||\vec{x}|| * ||\vec{w}|| * cos\theta$ we have that $p = \frac{<\vec{x},\vec{w}>}{||\vec{w}||}$
    - – If p>0 the hyperplane is in the direction of $\vec{w}$, if it is negative it is in the opposite direction
    - – Defining $b = -\frac{p}{||\vec{w}||}$ we have the canonical equation for the hyperplane
        - ∗ $w_1 x_1 + w_2 x_2 + b = 0$ in 2 dimensions
        - ∗ $< \vec{w}, \vec{x} > + b = W^t X + b = 0$ in n dimensions
    - – An hyperplane is useful for subdividing space
- 2 hyperplanes are parallel if their are defined by the same vector $\vec{w}$ allowing for a scaling factor λ
    - – $< Y, W > = \lambda < X, W >$
- The distance between parallel hyperplanes is computed as the difference of their projections on $\vec{w}$
    - – $d(X, Y) = p_y - p_x = \frac{b_x - b_y}{||W||}$
- The distance of a point A from an hyperplane is the projection of the point on the defining vector $\vec{w}$, minus the projection of the hyperplane on the same vector
    - – $D(A, X) = p_a - p_x = \frac{<A,W>+b}{||W||}$
- Hyperplanes are useful for the separation of classes of data
- Every column of a matrix can be thought of as a vector
    - – To make the dot product of 2 vectors using matrices you can multiply one vector for the transpose of the second
    - – $< \vec{a}, \vec{b} > = A * B^t$

Matrices

- A matrix is an array of numbers arranged in a rectangular structure
- The columns of a matrix are the coordinates where the basis vectors land after the transformation
- It has m rows and n columns, it is represented as $A \in R^{m*n}$
    - – $A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{,m2} & ... & a_{mn} \end{pmatrix}$
    - – The single a_{ij} numbers are called elements
    - – The index of an element is always mn, meaning first row and then column
- If n=1, the matrix is called column matrix, which is a vector
- If m=1, the matrix is a row matrix
- A and B are equal if they have the same dimensions and they are equal element by element
    - – $A = B \iff a_{ij} = b_{ij}$
- The 0 matrix contains all 0 elements and does not change the matrix it is added to
- The sum is defined as the sum of the respective elements
    - – We can sum only matrices of the same dimensions, they are said to be conformable for addition
    - – $C = B + A \iff c_{ij} = b_{ij} + a_{ij}$
    - – The difference operates in the same way
- Scalar multiplication is performed multiplying all the elements of the matrix for the scalar
    - – $C = \lambda A$ implies $c_{ij} = \lambda a_{ij}$
- The negative of A is -A, defined as $-1 * A$
    - – $A - A = 0$

- Matrix addition and scalar multiplication are commutative, associative and distributive
- Matrix product is an operation that is defined only if the number of rows of the first matrix is equal to the number of columns of the second (the matrices are conformable for the product)
  - A is of dimensions $m * p$ and B of dimensions $p * n$, if $C = A * B$
  - $c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj}$
  - $C = A * B$ can be computed as row by column product
  - It can be defined only if the number of columns in the first matrix is equal to the number of rows of the second
    * $R^{m*p} * R^{p*n} \implies R^{m*n}$
  - The result is a matrix with the same number of rows as the first, and the same number of columns as the second
  - The product between matrices is NOT commutative (!)
  - $A(B + C) = AB + AC$
  - $(A + B)C = AC + BC$
  - $A(BC) = (AB)C$
  - Be aware!
    * If $AB = 0$ we can NOT conclude that B or C are 0
    * If $AB = AC$ we can NOT conclude that $B = C$
- A square matrix has m=n
- An upper triangular matrix has all the elements below the diagonal equal to 0, and a lower triangular the ones above it
- A diagonal matrix has all the elements outside the diagonal equal to 0
- A diagonal matrix with all 1 elements is the identity matrix I
  - It does not change the square matrix it is multiplied to
  - In this case, $AI = IA = A$
- If AB=BA, A and B are said to commute
  - If A is a square matrix, it commutes with itself and with I
- If AB=-BA, A and B are said to anti-commute
- The transposition of a $n * m$ matrix is a $m * n$ matrix, called A^t, where $[A^t]_{ij} = A_{ji}$
  - A and A^t are always conformable to product, in both directions
  - $(A^t)^t = A$
- A square matrix is symmetric if $A = A^t$, antisymmetric (skew-symmetric) if $A = -A^t$
  - $A + A^t$ is always symmetric
  - $A - A^t$ is always antisymmetric
  - An antisymmetric matrix has a 0 diagonal and antisymmetrical elements otherwise
- The inverse of a matrix A, called A^{-1}, is a matrix such that $A * A^{-1} = A^{-1} * A = I$
  - It is defined only if $det(A) \neq 0$
- An orthogonal matrix has its inverse equal to the transpose, $A^{-1} = A^t$
  - An orthogonal matrix describes a spatial rotation
  - Therefore, $AA^t = A^t A = I$
    * You can check for orthogonality by checking that $A * A^t = I$
- Some properties of transpose and inverse matrices
  - $(AB)^{-1} = B^{-1} * A^{-1}$, but only if $(AB)^{-1}$ exists(!)
  - $(AB)^t = B^t * A^t$
- It is possible to associate a number called determinant to any square matrix
  - $det(A) = |A| \in R$
  - For an order 2 square matrix, that is compute subtracting the product of the second diagonal to that of the first
    * $det(A) = a_{11} * a_{22} - a_{12} * a_{21}$
  - It represents the area of the unit square after the transformation
  - Its sign reflects the orientation of space
    * If it is negative, the transformation flips the axis
- The rank of a transformation is the dimensionality of its output space, called column space
  - The column space of a transformation is the span of the basis vectors defined by its columns

- Some proprieties of determinants
  - If an entire row or column is equal to 0, then the determinant of the matrix is 0
  - $det(A * B) = det(A) * det(B)$
  - The determinant of an orthogonal matrix is either 1 or -1
  - $det(A) = det(A^t)$
- How to compute the inverse of 2*2 matrices
  - Given the definition $A * A^{-1} = I$ if $det(A) \neq 0$
  - It follows $A^{-1} = \frac{1}{det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$
- A minor M_{ij} of a matrix A is the determinant of any square submatrix of A
- The cofactor of the element a_{ij} is $C_{ij} : C_{ij} = M_{ij} * (-1)^{i+j}$
- To compute the determinant of any matrix you pick any row or column and sum the product of any element in it for its cofactor
  - In a column $det(A) = \sum\limits_{i=1}^{n} a_{ij} * C_{ij}$
  - In a row $det(A) = \sum\limits_{j=1}^{n} a_{ij} * C_{ij}$
  - It is convenient to choose the row or column with most 0 for the computation
  - If 2 rows are identical, det(A)=0
  - If one row is 0, then det(A)=0
  - If you exchange 2 rows, det(A')=-det(A)
  - The determinant of a triangular matrix is the product of the diagonal elements
  - If B is obtained by multiplying every element in a row of A by λ, $det(B) = \lambda det(A)$
  - For any n*n square matrix, $det(\lambda A) = \lambda^n det(A)$
  - If A and B are of the same order, $det(AB) = det(A)det(B)$
- The cofactor matrix of A, called A^c, is a matrix with each element equal to the cofactor of the same element in a
  - $A^c : a_{ij}^c = C_{ij}$
- The adjugate matrix of A is the transpose of its cofactor matrix
  - $A^a = (A^c)^t$
- The inverse matrix can be obtained by dividing the adjugate of a matrix for its determinant
  - $A^{-1} = \frac{1}{|A|} A^a$
- Matrices can represent systems of linear equations
  - The system $\begin{cases} x + y = 7 \\ 3x - y = 5 \end{cases}$ can be represented as $\begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$
  - The system has a solution if the coefficient matrix is invertible

Linear transformations

- A matrix can be thought of as a linear transformation of a vector space
  - $A^{m*n} : R^n \rightarrow R^m$
  - A linear transformation is a transformation that preserves linearity and does not move the origin
    * $A(\vec{v} + \vec{u}) = A\vec{v} + A\vec{u}$ and $A(\lambda\vec{v}) = \lambda A\vec{v}$
- A rotation by an angle θ can be describe by the transformation
  - $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
- The inverse trasformation takes a transformed vector and restores the original one
  - Sometimes it does not exist (!), when det(A)=0
- If det(A)=0 the transformation squishes space to a lower dimensional vector space
- The composition of the transformations A followed by B is $C = BA \neq AB$
- The scalar product of the transformation of a vector $A\vec{v}$ and the vector $\vec{w}$ is equal to the scalar product of the firts vector with the second vector transformed by the transpose of A
  - $A\vec{v} * \vec{w} = \vec{v} * A^t\vec{w}$

- The null space of a transformation is the set of vectros that get squished to $\vec{0}$ by the trasformation
  - $\vec{b} \in Null(A) \iff A\vec{b} = \vec{0}$
  - A trivial null space is always $\vec{0}$ itself
  - There is a true null space only if $det(A) = 0$
  - If $det(A) \neq 0$, the only null space is $\vec{0}$ itself
- The null space of a square matrix can be computed setting up a linear system of equations
  - For a matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, $det(A) = 0$
  - $A\vec{b} = \vec{0} \implies \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{cases} b_1 + 2b_2 = 0 \\ 2b_1 + 4b_2 = 0 \end{cases} \implies b_1 = -2b_2 \implies \vec{b} = \lambda \begin{pmatrix} -2 \\ 1 \end{pmatrix}$
- For the transformation A, if $A\vec{b} = \lambda\vec{b}$, $\vec{b}$ is an eigenvector of A and λ is its eigenvalue
  - An eigenvector of a transformation is a vector that is only rescaled by the transformation
  - An eigenvalue is the scaling factor to which the vector is subjected by the transformation
  - The $\vec{0}$ vector can never be an eigenvector even though $A\vec{0} = \lambda\vec{0}$ is always true
    * On the contrary, it is possible that $\lambda = 0$
- How to find eigenvalues for the matrix A
  - $A\vec{b} = \lambda\vec{b} \implies A\vec{b} - \lambda\vec{b} = 0 \implies (A - \lambda I)\vec{b} = 0$
  - This means that the eigenvectors $\vec{b}$ are the non-trivial null space of the transformation $(A - \lambda I)$
  - It is required that $det(A - \lambda I) = 0$, otherwise there are no eignevectors
  - The equation $det(A - \lambda I) = 0$ is called characteristic equation of A and its root allows to recover the eigenvalues of the matrix
- How to find the eigenvectors for the matrix A
  - Once I have the eigenvalues λ, the eigenvectors can be found by solving $(A - \lambda I)\vec{b} = 0$ for $\vec{b}$, using all the λ
- The number of eigenvalues and eigenvectors (families of linearly dependent eigenvectors) is equal to the dimensions of the vector space
  - I always have n families of eigenvectors in the vector space R^n
  - The families of eigenvectors are orthogonal to each other iff the transformation is symmetric
  - They define a convenient reference frame, even if they are not orthogonal
  - Each eigenvalue scales one of the families of eigenvectors, meaning that it stretches one of the dimensions of the new reference frame
- Note that in a trinagular or diagonal matrix the diagonal elements are its eigenvalues
- The product of the eigenvalues is equal to the determinant of the matrix
  - A non invertible matrix (also called singular matrix) has always at least a 0 eigenvalue
  - The inverse matrix has reciprocal eigenvalues $\left(\frac{1}{\lambda}\right)$
  - The eigenvalue of kA is $k\lambda$
  - The eigenvalue of A^n is λ^n
  - Transposition does not change the eigenvalues
- The sum of the eigenvalues is the trace of the matrix
  - The trace of a matrix is the sum of its diagonal elements
- Sometimes there can be no real eigenvalues, but there are always solutions in the complex field
  - This happens when the characteristic equation of the matrix has $\Delta < 0$
- A complex number z is written as $z = a + ib$ where $i = \sqrt{-1}$
  - a is called real part
  - b is called imaginary part
- The reference axis of a vector space are NOT necessarily orthogonal to each other (!)
  - But they must be linearly independent
- To go from one system to the other we need the representation of the new basis vectors in term of the previous ones
  - $\hat{i}' = a\hat{i} + b\hat{j}$ and $\hat{j}' = a\hat{i} + b\hat{j}$
  - If $U = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ we have that $\vec{v}' = U\vec{v}$
  - It is possible to go back to the original system of reference using U^{-1}

- If a $n*n$ matrix A has n eigenvectors which are linarly independent, I can write the $n*n$ matrix U containing all the eigenvectors, and use it to convert to a new system of reference
  - The eigenvectors of A will be the new basis vectors
- I can compute A in the new reference frame forming $\Lambda = U^{-1}AU$
  - Given a vector $\vec{v}$, I first convert it to the new reference frame where the eigenvectors are the basis vectors using U, then I apply A and finally I go back to the old system of reference using U^{-1}
  - This new matrix Λ will be diagonal (!)
  - Each column will be made of one eignevector multiplied by its eigenvalue
  - It is good to choose normalized vectors for the change of basis, meaning that their norm should be 1
    * In this case $det(U) = 1$, meaning that areas are preserved by the trasformation
- Why do I want to use eigenvectors as reference frames?
  - Because the components of any vector are only rescaled by the original transformation A in this reference frame
  - This makes much easier to compute transformations
- If a matrix is symmetric $(A = A^t)$ its eigenvalues are real and its eigenvectors are orthogonal
  - If the eigenvectors are normalized $U^{-1} = U^t$, therefore $\Lambda = U^t AU$
- In the same way that a linear form can be represented as all the points orthogonal to a vector with a projection p onto it, a matrix can describe a quadratic form
- A quadratic form is an equation in more than 1 variable were each term has a variable squared or multiplied to another variable
  - An example is $ax^2 + bxy + cy^2 = 0$
- This is represented as $\vec{x}^t A\vec{x} + b = 0$, where $\vec{x}$ is the vector containing the variables, and A is a matrix of coefficients
  - The vector is multiplied 2 times to reflec the fact that the expression is quadratic
  - The second time the transpose is used in order to allow the product
- By rescaling A, we can obtain the standard form $\vec{x}^t A\vec{x} = 1$
- The matrix that describes a quadratic form is always symmetric
  - If it is not singular (non-invertible), it can be diagonalised as $\Lambda = U^t AU$
  - If $\vec{x'} = U^t \vec{x}$, the quadratic form becomes $\vec{x'}^t \Lambda \vec{x'}$, defined canonical form
- In the canonical form, a $2*2$ Λ contains the eigenvalues of the transformation in the diagonal
  - If they are both positive, the quadratic is an ellipse
    * If they are equal, it is a circle
  - If they are of opposite sign, it is an hyperbole
  - If they are both negative, there is no real solution
- In 3d, I can get an ellipsoid, a hyperboloid of 1 sheet or an hyperboloid of 2 sheets

---

rewied until here

25/10/19

One variable calculus

- Calculus is the study of functions
  - Functions are univocal relations between the sets domain and codomain
- The function $f(x) = mx + q$ is a line passing through q at $x = 0$ with slope m
- The inverse of a function correlates $f(x)$ to x
  - It is the reflecion of $f(x)$ on the line $g(x) = x$
- The function $f(x) = a^x$ is an exponential
  - It passes through 1 at $x = 0$
  - $\lim_{x \to -\infty} f(x) = 0$
  - $\lim_{x \to +\infty} f(x) = +\infty$

- The function $f(x) = \log_a(x)$ is a logarthmic function
  - It passes through 1 at $x = 0$
  - Common bases a are 10, 2 and e
  - $\lim_{x \to 0} f(x) = -\infty$
  - $\lim_{x \to +\infty} f(x) = +\infty$
  - Logarithms are useful for performing products
  - $\log_a(xy) = \log_a(x) + \log_a(y)$
  - $\log_a(x^y) = y \log_a(x)$
  - $\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$
    * check this
- Trigonometric functions
  - The cosine is an even function beacuse $\cos(\theta) = \cos(-\theta)$
    * $\cos(\frac{\pi}{2}) = 0$
    * $\cos(0) = 1$
  - The sine is an odd function because $\sin(\theta) = -\sin(-\theta)$
  - Sine and cosine are periodical with a 2π period
  - Trigonometric functions can be inverted only in a subdomain
  - They are continous functions
- The derivative of a function is another function that describes its rate of change
  - $\frac{d}{dx} = f'(x)$
  - $f'(x)_{x=a} = \lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x}$
  - $\frac{d}{dx} ax = a$
  - $\frac{d}{dx} x^n = nx^{n-1}$
  - $\frac{d}{dx} k * f(x) = kf'(x)$
  - $\frac{d}{dx} g(x) * f(x) = g(x) * f'(x) + g'(x) * f(x)$
  - $\frac{d}{dx} \frac{g(x)}{f(x)} = \frac{g(x) * f'(x) + g'(x) * f(x)}{f(x)^2}$
  - $\frac{d}{dx} \cos(x) = -sin(x)$
  - $\frac{d}{dx} \sin(x) = cos(x)$
  - $\frac{d}{dx} e^x = e^x$
  - $\frac{d}{dx} \ln(x) = \frac{1}{x}$
  - There are functions that are continuous but not derivable
- There are global and local maxima and minima
  - There are NOT methods to compute global maxima and minima
- A local maximum/minimum has several proprieties
  - The derivative at that point is 0
    * This is NOT sufficient, it can also be a flexus
  - A minimum has a derivative with positive slope when it intersect the x axis
    * In other words, the second derivative has to be positive
  - A maximun has a derivative with negative slope when it intersect the x axis
    * In other words, the second derivative has to be negative

Taylor series

- A line passing in x_a with a slope equal to the derivative at that point approximates the function itself
  - $f(x) - f(x_a) \approx \frac{df}{dx}|_{x_a}(x - x_a)$
- We can approximate a function with a polinomial $P(x)$
- The coefficient a_0 for grade 0 can be found solving
  - $P(x_a) = f(x_a)$
- We can find the other terms by equating the second, third and fourth derivatives
  - $P'(x_a) = f'(x_a)$
  - $P''(x_a) = f''(x_a)$
  - $P'''(x_a) = f'''(x_a)$

- $P''''(x_a) = f''''(x_a)$
- In general, we can give the Taylor series
  - $P(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ...$

Integrals

- The area under a curve over a partition is called integral
  - This can be computed doing a Riemann sum, meaning that we can sum the area of rectangles under the curve
  - $F(x) = \int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{k=1}^{n}(f(x)*\Delta x)$
- The areas computed by integrals have a sign (!)
- It is possible to solve an integral by substituting a polinomial with the variable u
  - du then becomes $\frac{d}{dx}u\,dx$

Multi-dimensional calculus

- A 2 dimensional function takes 2 inputs x,y and gives the output z
  - $z = f(x,y)$
  - They are usually represented with 3d surfaces or with level curves
- It is not possible to compute single derivatives of the function
- We can fix y and compute the derivative with respect to x in the resulting 2d curve, called partial derivative in x
  - $\frac{\partial f}{\partial x} = f_x(x,y) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x,y)-f(x,y)}{\Delta x}$
- It is possible to take second partial derivatives an mixed partial derivatives
- In mixed partial derivatives the order of derivation is not important (!)
- The set of first partial derivatives is called gradient of the function
  - $\nabla f(x_1, x_2, ...x_n) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ...\frac{\partial f}{\partial x_n})$
- The square symmetric matrix of all second partial derivatives is called Hessian (H) of $f(x)$
- The minima and maxima of a multivariable function are those points where the gradient of the function is 0
- A minimum has the diagonal elements of the Hessian both positive, a maximum both negative
- If the diagonal of the Hessian has negative and positive elements we have saddle points
- When the Hessian is not diagonal, it can be diagonalized
- The maxima and minima of a multivariable function $f(x,y)$ subjected to the costraint $g(x,y) = 0$ are the points where $g(x,y)$ is tangent to the contour level of $f(x,y)$
  - We can also say that the perpendiculars to both functions are parallel
  - The perpendicular to the contour level is the gradient of the function
  - $\nabla f = \lambda \nabla g$ where λ is the Lagrange multiplier
- The Lagrangian function $L(\vec{(x)}, \lambda)$ is bad
- Information entropy

Probability

- Probability can be interpreted in several ways
  - In a frequency approach, it is the ratio between favourable events and total events
  - It can also be viewed as the confidence in an event happening
- The sample space is the set of all possible outcomes of an experiment
- An event is a subset of the sample space
- A probability is a number between 0 and 1
  - $0 \leq P(E_i) \leq 1$
- The sum of the probabilities of all the outcomes is equal to 1
  - $P(\cup E_i) = 1$

- The probability of on of 2 events happening is equal to the sum of the probabilities minus their intersection
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If two events have NO intersection they are said mutually exclusive
- The intersection of 2 events that are independent and not mutually exclusive is the product of their probabilities
  - $P(A \cap B) = P(A) * P(B)$
- The conditional probability of A given B is represented as $P(A|B)$
  - If the events are independent, $P(A|B) = P(A)$
  - If they are not independent, $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- The intersection of 2 non-independent event is therefore
  - $P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$
- 2 mutually exclusive events cannot be independent, and vice-versa
- The bayes formula: $P(A|B)$ is different from $P(B|A)$
  - $P(B|A) = \frac{P(A|B)*P(B)}{P(A)}$
- To count events considering order we use permutations
  - With replacement (same element can be drawn twice)
    * It is the number of events multiplied by itself for a number of times equal to the lenght of the sequence
    * l : sequence length, n : possible outcomes $\implies x = n^l$
  - Without replacement (same element cannot be drawn twice)
    * It is the same of the replacement case, but n becomes n-1 at each iteration
    * l : sequence length, n : possible outcomes $\implies x = \frac{n!}{(n-l)!}$
- To count events NOT considering order we use combinations
  - With replacement (same element can be drawn twice)
  - Without replacement (same element cannot be drawn twice)