What is Data Mining?

Data mining is the discovery of models for data

- **Statistical Modeling: Construction of a statistical model for the data**
- **Machine Learning (training and test datasets)**
- **Computational (Summarizing or Extracting prominent features)**

Big Data characteristics (five Vs)

- **Volume**

The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.

- **Variety**

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

- **Velocity**

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

- **Variability**

Inconsistency of the data set can hamper processes to handle and manage it.

- **Veracity**

The data quality of captured data can vary greatly, affecting the accurate analysis.

Linear Algebra

Review

Source: Linear Algebra and Matrices - UCL

Review Linear Algebra

- Definitions-Scalars
- Vectors and Matrices
- Vector and Matrix calculations
- Identity, inverse matrices & determinants
- Eigenvectors & inner products

Scaler

- A quantity (*variable*), described by a single real number

Vector

- **Series of numbers (e.g.** . A column of numbers) $\begin{bmatrix} x1 \\ x2 \\ xn \end{bmatrix}$

Not a physics vector (magnitude, direction)

Matrix

- Rectangular display of vectors in rows and columns

- Can inform about the same vector intensity at different

  times simultaneously .

- Vector is just a n x 1 matrix

$$\begin{bmatrix} x11 & x12 & x13 \\ x21 & x22 & x23 \\ x31 & x32 & x33 \end{bmatrix}$$

Transposition

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \mathbf{b}^T = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 3 & 4 & 9 \end{bmatrix} \qquad \mathbf{d}^T = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}$$

column $\longrightarrow$ row row $\longrightarrow$ column

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 4 & 1 \\ 6 & 7 & 4 \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 4 & 7 \\ 3 & 1 & 4 \end{bmatrix}$$

Matrix Calculations

Addition

- Commutative: **A**+**B**=**B**+**A**
- Associative: (**A**+**B**)+**C**=**A**+(**B**+**C**)

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}2 & 4\\ 2 & 5\end{bmatrix}+\begin{bmatrix}1 & 0\\ 3 & 1\end{bmatrix}=\begin{bmatrix}2+1 & 4+0\\ 2+3 & 5+1\end{bmatrix}=\begin{bmatrix}3 & 4\\ 5 & 6\end{bmatrix}$$

Subtraction

- By adding a negative matrix

$$\mathbf{A}-\mathbf{B}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}-\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}+\begin{bmatrix}-1 & -2\\ -3 & -4\end{bmatrix}=\begin{bmatrix}1 & 2\\ 2 & -1\end{bmatrix}$$

Scalar multiplication

Scalar * matrix = scalar multiplication

$$\lambda \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b & \lambda c \\ \lambda d & \lambda e & \lambda f \end{pmatrix}$$

Matrix Multiplication

$$AB = \underset{2\times 3}{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}} \cdot \underset{3\times 2}{\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}} = \underset{2\times 2}{\begin{pmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} \end{pmatrix}}$$

Matrix Multiplication

$$
\begin{array}{cc}
 & n \\
m & \begin{bmatrix} A_1 & A_2 & A_3 \\ A_4 & A_5 & A_6 \\ A_7 & A_8 & A_9 \\ A_{10} & A_{11} & A_{12} \end{bmatrix}
\end{array}
\quad
\begin{array}{cc}
l & \\
\begin{bmatrix} B_{13} & B_{14} \\ B_{15} & B_{16} \\ B_{17} & B_{18} \end{bmatrix} & k
\end{array}
\quad = \mathbf{m} \times \mathbf{l} \text{ matrix}
$$

Matrix multiplication

- Matrix multiplication is NOT commutative i.e the order matters!

  $AB \neq BA$

- Matrix multiplication IS associative

  $A(BC)=(AB)C$

- Matrix multiplication IS distributive

  $A(B+C)=AB+AC$

  $(A+B)C=AC+BC$

Identify Matrix

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

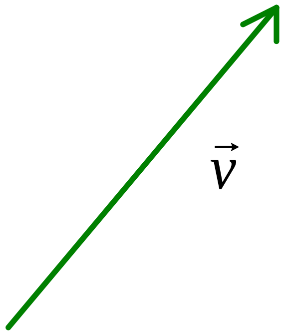
nxn matrix $\boldsymbol{A}$, we have $\boldsymbol{A}\,\boldsymbol{I}_n = \boldsymbol{I}_n\,\boldsymbol{A} = \boldsymbol{A}$

Part II

More Advanced Matrix Techniques

Vector components & orthonormal base

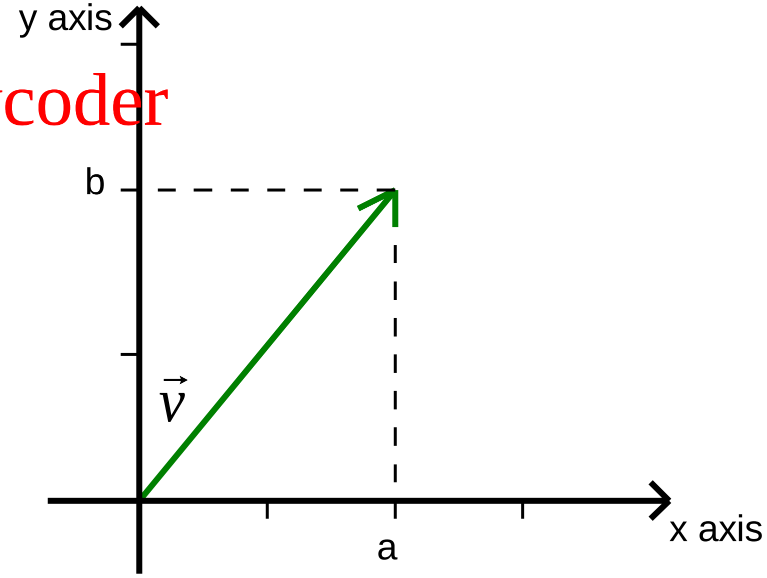
- A given vector (a b) can be summarized by its *components*, but only in a particular *base* (set of axes; the vector itself can be independent from the choice of this particular base).



example

a and b are the components of $\vec{v}$

in the given base (axes chosen for expression of the coordinates in vector space)



Orthonormal base: *set of vectors chosen to express the components of the others, perpendicular to each other and all with norm (length) = 1*

Linear combination & dimensionality

Vectorial space: space defined by different vectors (for example for dimensions…).
The vectorial space defined by some vectors is a space that contains them and all the vectors that can be obtained by multiplying these vectors by a real number then adding them (*linear combination*).

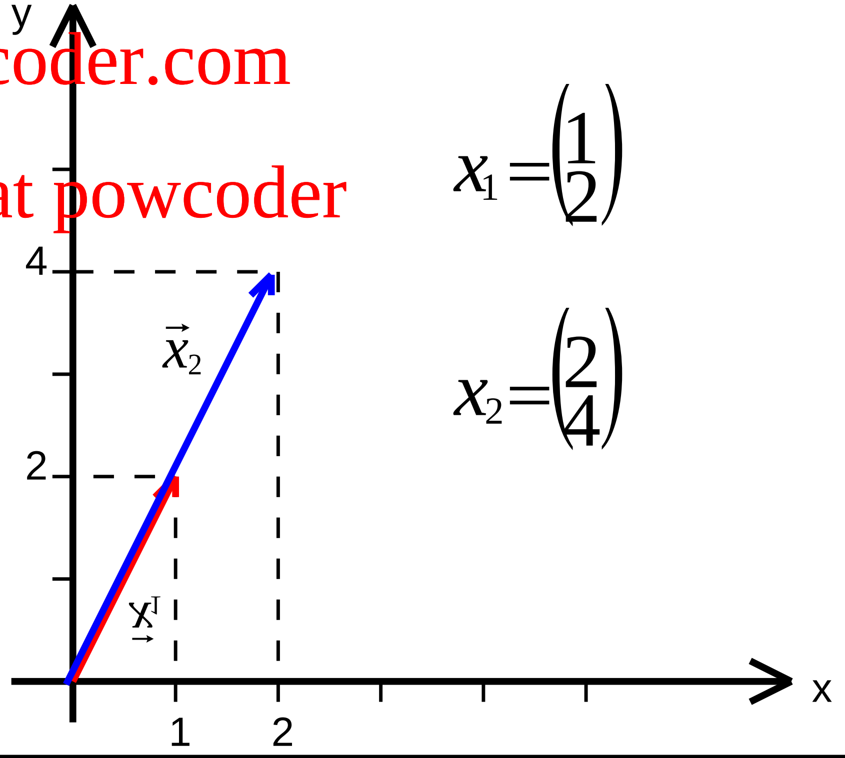
A matrix A ($m \times n$) can itself be decomposed in as many vectors as its number of columns (or lines). When decomposed, one can represent each column of the matrix by a vector. The ensemble of n vector-column defines a *vectorial space* proper to matrix A.
Similarly, A can be viewed as a matricial representation of this ensemble of vectors, expressing their components in a given *base*.

Linear dependency and rank

If one can find a *linear relationship* between the lines or columns of a matrix, then the *rank* of the matrix (number of dimensions of its vectorial space) will not be equal to its number of column/lines – the matrix will be said to be *rank-deficient*.

Example

$$X = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$



$$x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

When representing the vectors, we see that x1 and x2 are superimposed. If we look better, we see that we can express one by a *linear combination* of the other: x2 = 2 x1.

The *rank* of the matrix will be 1.
In parallel, the *vectorial space* defined will has only one dimension.

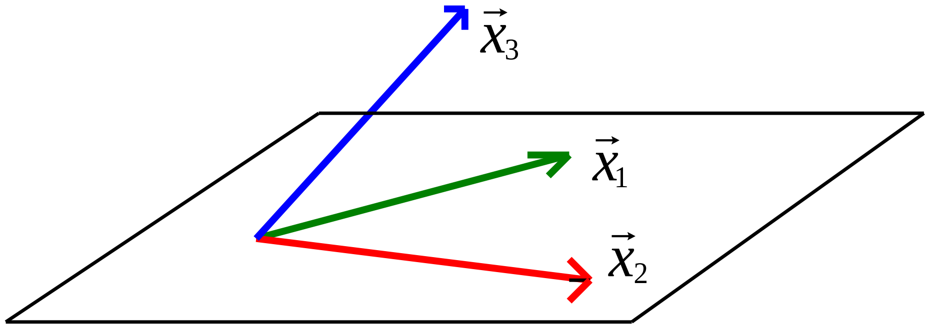
Linear dependency and rank

• The *rank of a matrix* corresponds to the *dimensionality* of the vectorial space defined by this matrix. It corresponds to the number of vectors defined by the matrix that are linearly independents from each other.

• *Linealy independent* vectors are vectors defining each one one more dimension in space, compared to the space defined by the other vectors. They cannot be expressed by a linear combination of the others.

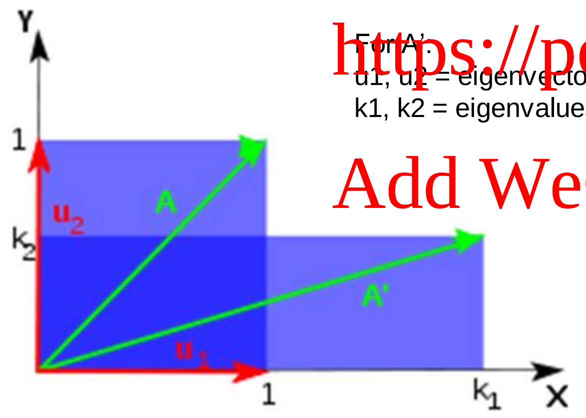
Note. *Linearly independent* vectors are not necessarily *orthogonal* (perpendicular).
Example: take 3 linearly independent vectors x1, x2 et x3.

Vectors x1 and x2 define a plane (x,y)
And vector x3 has an additional non-zero component in the z axis.
But x3 is not perpendicular to x1 or x2.

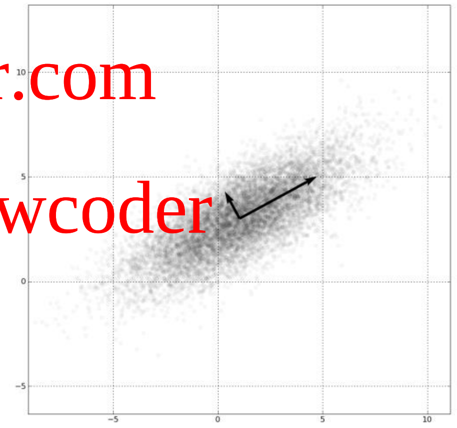


Eigenvalues et eigenvectors

One can represent the vectors from matrix *X* (*eigenvectors* of A) as a set of *orthogonal vectors* (perpendicular), and thus representing the different dimensions of the original matrix A.
The *amplitude* of the matrix *A* in these different dimensions will be given by the *eigenvalues* corresponding to the different *eigenvectors* of A (the vectors composing X).
Note: if a matrix is rank-deficient, at least one of its eigenvalues is zero.



For A':
u1, u2 = eigenvectors
k1, k2 = eigenvalues



In *Principal Component Analysis (PCA)*, the matrix is decomposed into *eigenvectors* and *eigenvalues* AND the matrix is *rotated* to a new *coordinate system* such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first *principal component*), the second greatest variance on the second coordinate, and so on.

Vector Products

Two vectors:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Inner product = scalar

Inner product X^TY is a scalar
(1xn) (nx1)

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1y_1 + x_2y_2 + x_3y_3 = \sum_{i=1}^{3} x_i y_i$$

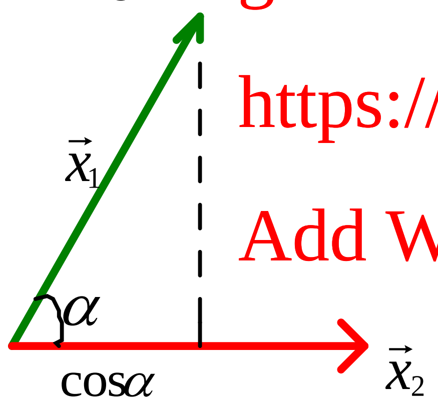
Outer product = matrix

$$\mathrm{xy}^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{bmatrix}$$

Outer product XY^T is a matrix
(nx1) (1xn)

Scalar product of vectors

Calculate the *scalar product* of two vectors is equivqlent to make the *projection* of one vector on the other one. One can indeed show that $\vec{x1} \bullet \vec{x2} = |\vec{x1}| \, . \, |\vec{x2}| \, . \, \cos\alpha$ *where* α is the angle that separates two vectors when they have both the same origin.



$$x1 \bullet x2 = |\vec{x}_1| \, . \, |\vec{x}_2| \, . \, \cos\alpha$$

In parallel, if two vectors are orthogonal, their scalar product is zero: the projection of one onto the other will be zero.

Determinants

For a matrix 1×1:

$$\det(a_{11}) = a_{11}$$

For a matrix 2×2:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

For a matrix 3×3:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

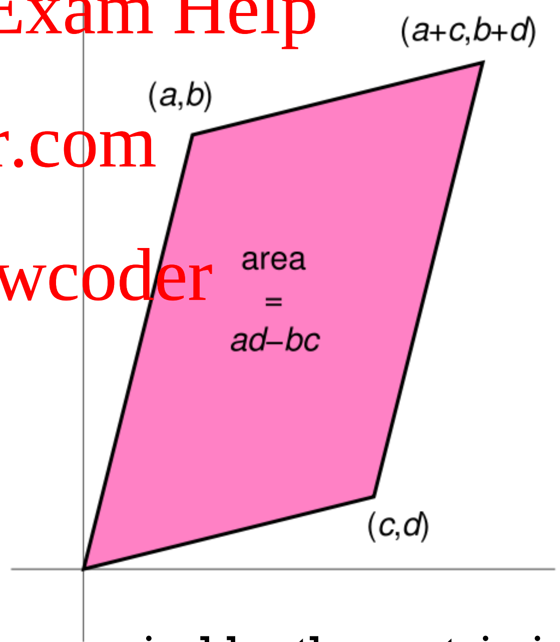
The determinant of a matrix can be *calculate by multiplying* each *element* of one of its lines by the *determinant of a sub-matrix* formed by the elements that stay when one suppress the line and column containing this element. One give to the obtained product the sign $(-1)^{i+j}$.

Determinants

•Determinants can only be found for square matrices.

•For a 2x2 matrix A, **det(A) = ad-bc.** Lets have at closer look at that:

$$\det(A) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$



The determinant gives an idea of the 'volume' occupied by the matrix in vector space

A matrix A has an inverse matrix A^{-1} if and only if **det(A)≠0.**

Determinants

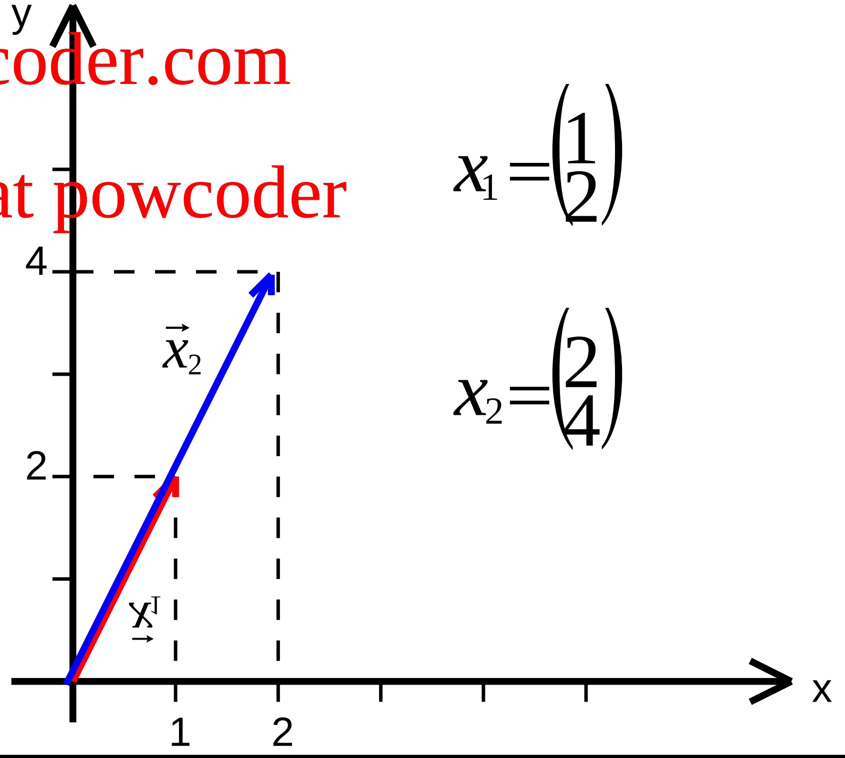
The *determinant* of a matrix is *zero* if and only if there exist a linear relationship between the lines or the columns of the matrix – if the matrix is *rank-deficient*. In parallel, one can define the *rank* of a matrix A as the size of the largest square sub-matrix of A that has a non-zero determionant.

$$X = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

Here x1 and x2 are superimposed in space, because one can be expressed by a *linear combination* of the other: x2 = 2 x1.

The *determinant* of the matrix X will thus be zero.

The largest square sub-matrix with a non-zero determinant will be a matrix of 1x1 => the *rank* of the matrix is 1.



$$x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

Determinants

- In a vectorial space of n dimensions, there will be no more than n linearly independent vectors.
- If 3 vectors (2×1) x'_1, x'_2, x'_3 are represented by a matrix X':

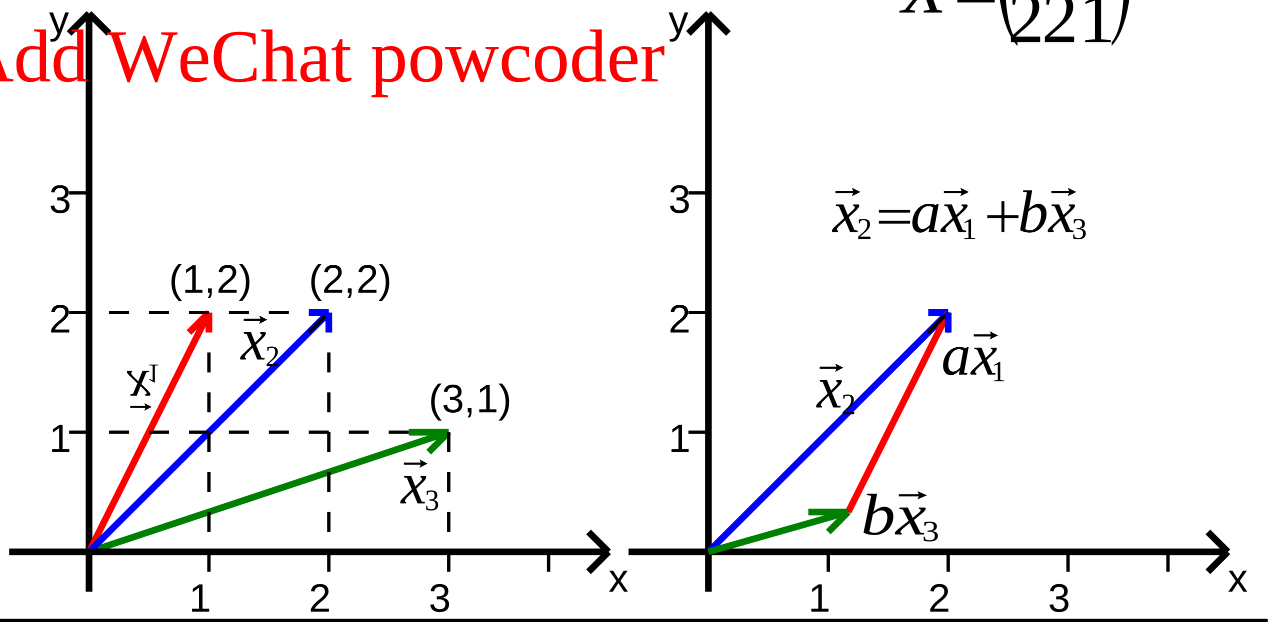
$$X' = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \end{pmatrix}$$

Graphically, we have:

Here x3 can be expressed by a *linear combination* of x1 and x2.

The *determinant* of the matrix X' will thus be zero.

The largest square sub-matrix with a non-zero determinant will be a matrix of 2x2 => the *rank* of the matrix is 2.



Determinants

The notions of *determinant*, of the *rank* of a matrix and of linear dependency are closely linked.

Take a set of vectors $x1, x2,\ldots,xn$, all with the same number of elements: these vectors are *linearly dependent* if one can find a set of scalars $c1, c2,\ldots,cn$ non equal to zero such as:

$$c1\ x1 + c2\ x2 + \ldots + cn\ xn = 0$$

A set of vectors are linearly dependent if one of then can be expressed as a *linear combination* of the others. They define in space a smaller number of dimensions than the total number of vectors in the set. The resulting matrix will be *rank-deficient* and the *determinant* will be *zero*.

Similarly, if all the elements of a line or column are zero, the determinant of the matrix will be zero.

If a matrix present two rows or columns that are equal, its determinant will also be zero

Matrix inverse

- **Definition.** A matrix ***A*** is called **nonsingular** or **invertible** if there exists a matrix ***B*** such that:

$$\boxed{A\,B = B\,A = I_n}$$

$$\begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3}+\frac{1}{3} & \frac{-1}{3}+\frac{1}{3} \\ \frac{-2}{3}+\frac{2}{3} & \frac{1}{3}+\frac{2}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- **Notation.** A common notation for the inverse of a matrix ***A*** is ***A***$^{-1}$. So:

$$\boxed{A\,A^{-1} = A^{-1}\,A = I_n\,.}$$

- The inverse matrix is unique when it exists. So if ***A*** is invertible, then ***A***$^{-1}$ is also invertible and then $(A^T)^{-1} = (A^{-1})^T$

- In Matlab: ***A***$^{-1}$ = **inv(*A*)**

- Matrix division: A/B= ***A**B***$^{-1}$

Matrix inverse

- For a XxX square matrix: $A = \begin{pmatrix} x_{1,1} & \dots & x_{1,j} \\ \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,j} \end{pmatrix}$

- The inverse matrix is: $A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} \operatorname{cof}(A, x_{1,1}) & \dots & \operatorname{cof}(A, x_{1,j}) \\ \vdots & \ddots & \vdots \\ \operatorname{cof}(A, x_{i,1}) & \dots & \operatorname{cof}(A, x_{i,j}) \end{pmatrix}^T$

- E.g.: 2x2 matrix $\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$

For a matrix to be invertible, its determinant has to be non-zero (it has to be square and of full rank).
A matrix that is not invertible is said to be *singular*.
Reciprocally, a matrix that is *invertible* is said to be *non-singular*.

Review

Review of Probability and Statistics

Probability in discrete space

Probability Axioms:

$$P(A) \geq 0$$

$$P(\Omega) = 1$$

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

A1	A2	A3	A4
A5	A6	A7	A8
A9	A10	A11	A12
A13	A14	A15	A16
A17	A18	A19	A20

Classical Probability

- **Assumes all outcomes in the sample space are equally likely to occur**

Classical probability of event A:

$$P(A) = \frac{N_A}{N} = \frac{\text{number of outcomes that satisfy the event A}}{\text{total number of outcomes in the sample space}}$$

- **Requires a count of the outcomes in the sample space**

Probability in discrete space

Lemma:

$$P(A) = 1 - P(\overline{A})$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Expectation of function variables

$$E[f(x)] = \sum_x f(x)p(x)$$

$$E(g(X)) = \int g(X)f(X)$$

Expectation of a random variables

$$\mu = E[X] = \sum_x xp(x)$$

$$E[X] = \int Xf(X)$$

E[a]=a; E[aX]=aE[X]

Covariance between two random variables

$$\mathrm{Cov}[x, y] = \sum\sum(x\text{-}\mu x)(y\text{-}\mu y)p(x,y)$$

$$\mathrm{Cov}[x, y] = \int\int(x - \mu x)(y - \mu y)f(x, y)$$

Variance of a random variables

$$Var(X) = E[(X-\mu)^2]$$

Var(a)=0; Var(aX)=a^2Var(X)