Week 8:
Least-Squares Data Fitting
Determinants

MSIN00180 Quantitative Methods for Business

In[ ]:=
```
$Post:=If[MatrixQ[#]||VectorQ[#],MatrixForm[#],#]&;
(*$Post:=If[MatrixQ[#],MatrixForm[#],#]&;*)
```

Transpose A^T

The **transpose** of a matrix A is the matrix A^T whose rows vectors are the same as the column vectors of A.

e.g.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Transposing Matrices in Mathematica

In[67]:=

```
Transpose[(1 4
           2 5
           3 6)]
```

$$\text{Transpose}\left[\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}\right]$$

Out[67]//MatrixForm=

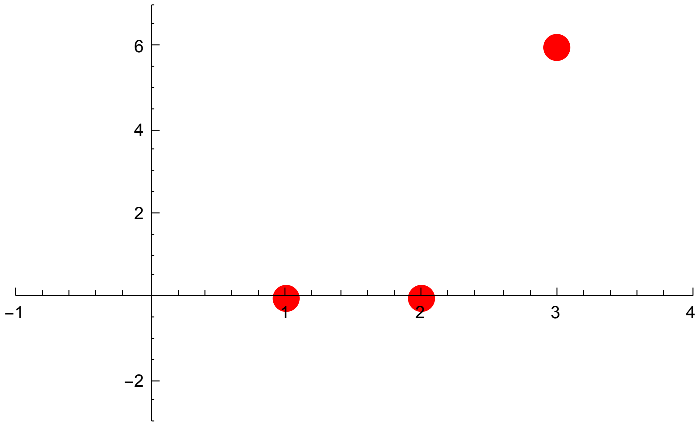
$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Some Properties of the Transpose

It can be shown that the following properties of transposed matrices holds true:

- $(A + B)^T = A^T + B^T$
- $(k\,A)^T = k\,A^T$
- $(A\,B)^T = B^T\,A^T$
- $\mathrm{rank}(A^T) = \mathrm{rank}(A)$
- $(A^T)^{-1} = (A^{-1})^T$

Least-squares data fitting



Problem: Fit a straight line equation $c_0 + c_1 x = y$ to the above 3 data points.

We can represent each data point as a separate equation:

$c_0 + c_1\, 1 = 0$ for point (1,0)
$c_0 + c_1\, 2 = 0$ for point (2,0)
$c_0 + c_1\, 3 = 6$ for point (3,6)

In matrix form $\Longrightarrow$ $\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$

Express the problem as $A\vec{x} = \vec{b}$

$$A\vec{x} = \vec{b} \quad \text{where } A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \text{ and } \vec{b} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$$

Naming the coefficient vector $\vec{x}$ is a bit confusing as it contains coefficients rather than "x" coordinate values.

- We do this to be consistent with the textbook and because $\vec{x}$ typically is used to represent the input vector of a linear transformation.

$\vec{b}$ contains the **actual output values** of our data points.

Would you expect $A\vec{x} = \vec{b}$ to have a solution?

Unless data points all lie on line being fitted $A\vec{x} = \vec{b}$ has no solution!

Unless the data points exactly lie on the type of line we are trying to fit the data to, the solution to $A\vec{x} = \vec{b}$ will be **inconsistent** - so

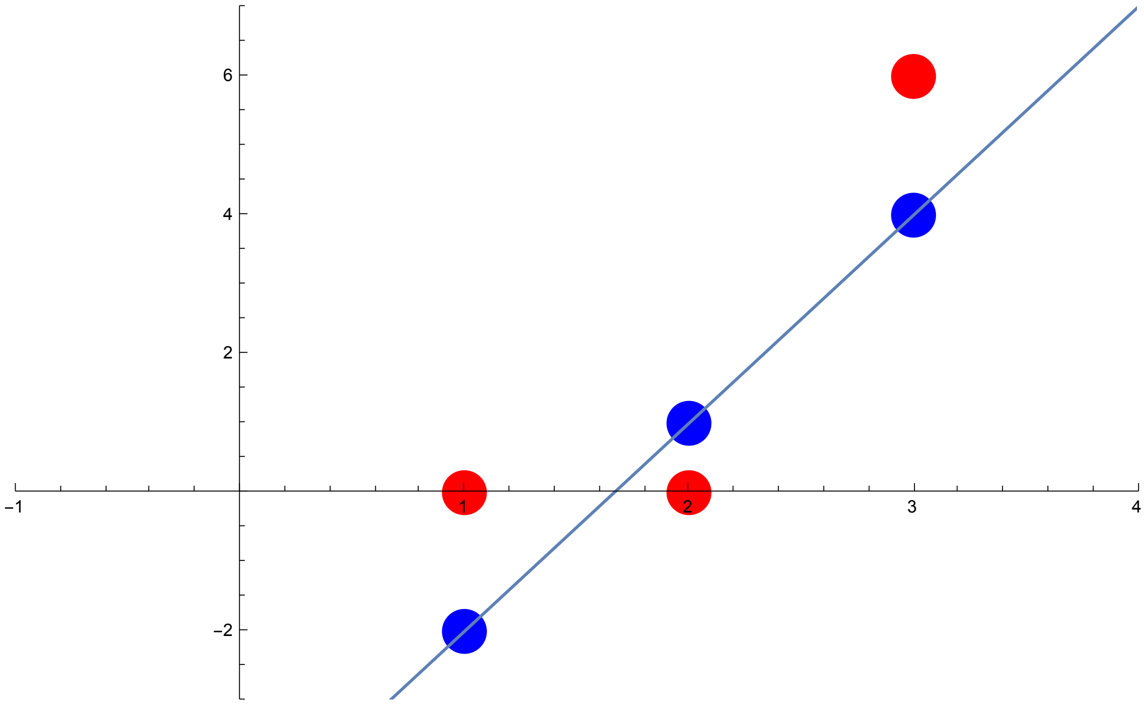
$$A\vec{x} \neq \vec{b}\,!$$

Whereas $\vec{b}$ contains the **actual output values** of our data points, the vector $A\vec{x}$ contains **predicted output values** based on the line specified by the coefficients in $\vec{x}$.

Try guessing ...

Consider if we **guessed** coefficients for a line we believed might be a good fit

$c_0 + c_1 x = y$ where $c_0 = -5$, $c_1 = 3$ $\implies$ $\vec{x} = \begin{pmatrix} -5 \\ 3 \end{pmatrix}$



The **predicted** output values are given by $A\,\vec{x} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} -5 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$

The **actual** output values $\vec{b} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$

Error between the actual and predicted output values

We can express the **error** between the **actual** output values ($\vec{b}$) and the **predicted** output values as **magnitude** of the difference $\vec{b} - A\,\vec{x}$:

error = $|| \vec{b} - A\,\vec{x} ||$

Error between the actual and predicted output values

For our example:

$$\vec{b} - A\,\vec{x} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} \quad = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}$$

The **magnitude** of this vector is given by

$$\|\vec{b} - A\,\vec{x}\| = \left\| \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} \right\| \quad = \sqrt{2^2 + (-1)^2 + 2^2} \quad = 3$$

This measure of overall error between a **prediction model** and a set of data points involves calculating the **sum of the squares** of the individual error terms.

Finding a "best fit" prediction model $\vec{x}^*$

We need to formulate a different problem that returns an approximate solution $A\vec{x}^*$ that is **"as close as possible"** to $\vec{b}$

In this case, "close as possible" means to find a solution $\vec{x}^*$ that **minimises** $\|\vec{b} - A\vec{x}^*\|$ with respect to all other possible solutions.

Least-squares solution

Consider a linear system $A\vec{x} = \vec{b}$ where A is an $n \times n$ matrix.

A vector $\vec{x}^*$ in $\mathbb{R}^n$ is called the **least-squares solution** of this system if

$$\|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\| \text{ for all } \vec{x} \text{ in } \mathbb{R}^n.$$

The term **least-squares solution** reflects the fact we are minimizing the **sum of the squares** of the components of the vector $\vec{b} - A\vec{x}$.

The linear system $A\vec{x} = \vec{b}$ is taken to be **inconsistent**.

Minimise $|| \vec{b} - A\vec{x} ||$

Minimising $|| \vec{b} - A\vec{x} ||$ is the same as minimising the sum of the squares of the components of $(\vec{b} - A\vec{x})$, which is given by the dot product:

$(\vec{b} - A\vec{x})^T.(\vec{b} - A\vec{x})$

Expanding this dot product we get:

$(\vec{b} - A\vec{x})^T.(\vec{b} - A\vec{x}) = (\vec{b}^T - \vec{x}^T A^T).(\vec{b} - A\vec{x})$ $\longleftarrow$ since $(A.B)^T = B^T A^T$

$= \vec{b}^T \vec{b} - \vec{x}^T A^T \vec{b} - \vec{b}^T A\vec{x} + \vec{x}^T A^T A\vec{x}$

To minimise this expression we need to employ **matrix differentiation** ...

Matrix differentiation

We are not going to formally cover or assess matrix differentiation in this module.

However, the following **matrix differentiation** rules are needed to differentiate our least-squares error expression:

Rule 1 : $\frac{\partial\,(A\,\vec{x})}{\partial\vec{x}} = A$

Rule 2 : $\frac{\partial\,(\vec{x}^T A)}{\partial\vec{x}} = A^T$

Rule 3 : $\frac{\partial\,(\vec{x}^T A^T A\,\vec{x})}{\partial\vec{x}} = 2\,\vec{x}^T A^T A$

While the first rule may seem reasonably familiar, the others are a little more obscure and require separate proofs that we are not going to cover.

Differentiate error expression and set to zero

Using the **matrix differentiation rules** from the previous slide we can now differentiate our **least-squares error expression**:

$$\frac{\partial}{\partial \vec{x}}\left(\vec{b} - A\,\vec{x}\right)^2 = \frac{\partial}{\partial \vec{x}}\left(\vec{b}^T\,\vec{b} - \vec{x}^T A^T\,\vec{b} - \vec{b}^T A\,\vec{x} + \vec{x}^T A^T A\,\vec{x}\right)$$

$$= -\frac{\partial}{\partial \vec{x}}\left(\vec{x}^T A^T\,\vec{b}\right) - \frac{\partial}{\partial \vec{x}}\left(\vec{b}^T A\,\vec{x}\right) + \frac{\partial}{\partial \vec{x}}\left(\vec{x}^T A^T A\,\vec{x}\right)$$

$$= -2\,\vec{b}^T A + 2\,\vec{x}^T A^T A$$

If we now **set this to zero** (for a minimum) we get:

$$\vec{b}^T A = \vec{x}^T A^T A$$

Now take the **transpose** of both sides:

$$\left(\vec{b}^T A\right)^T = \left(\vec{x}^T A^T A\right)^T$$

$\boldsymbol{A^T\,\vec{b} = A^T\,A\,\vec{x}}$ ⟵ This is called the **Normal Equation**.

The Normal Equation

We have now shown that ...

The **least-squares solutions** of the system $A\vec{x} = \vec{b}$ are the exact solutions of the (consistent) system

$$A^T A \vec{x}^* = A^T \vec{b}$$

The system $A^T A \vec{x}^* = A^T \vec{b}$ is called the **normal equation** of $A\vec{x} = \vec{b}$.

Alternative form of the normal equation

Provided the matrix $(A^T A)$ is invertible, a **direct formula for the least-squares solution $\vec{x}^*$** to the system $A\vec{x} = \vec{b}$ is given by

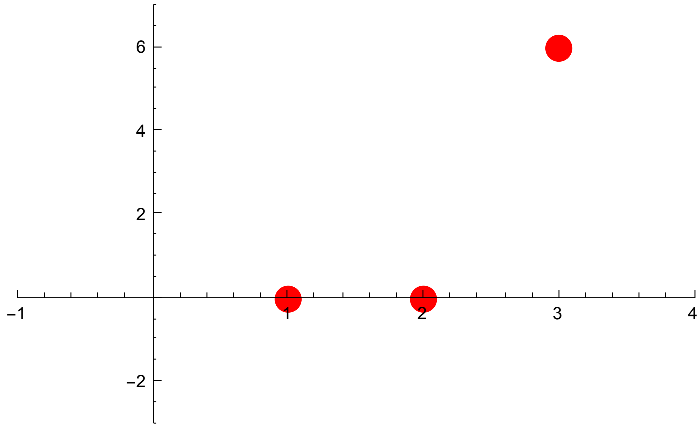
$$\vec{x}^* = (A^T A)^{-1} A^T \vec{b}$$

Note:

It is normally easier to solve the **normal equation** $A^T A \vec{x}^* = A^T \vec{b}$ by **Gauss-Jordan elimination** than by calculating the inverse of $A^T A$ to use the above formula.

Example

Consider our earlier example ...



Remember we can express the above problem as the system:

$A\vec{x} = \vec{b}$ where $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$, $\vec{x} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$

where c_0, c_1 are the coefficients of a straight line equation given by $c_0 + c_1 x = y$

Example

Consider the **normal equation**

$$A^T A \vec{x}^* = A^T \vec{b}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \vec{x}^* = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix} \vec{x}^* = \begin{pmatrix} 6 \\ 18 \end{pmatrix}$$

Now form this into an augmented matrix and use **Gauss-Jordan elimination** ...

Example 1

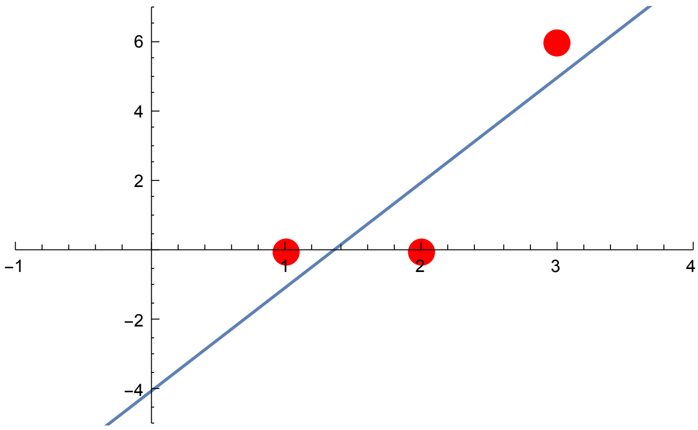
In[87]:=
```
m=(3  6   6
   6  14  18);
m⟦1⟧ /= 3;
m⟦2⟧ -= 6 m⟦1⟧;
m⟦2⟧ /= 2;
m⟦1⟧ -= 2 m⟦2⟧;
m
```

Out[92]//MatrixForm=

$$\begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 3 \end{pmatrix}$$

$$\text{RREF} = \begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 3 \end{pmatrix} \Longrightarrow \vec{x}^* = \begin{pmatrix} -4 \\ 3 \end{pmatrix} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$$

The least-squares best fit line is given by the line $\boxed{-4 + 3x = y}$



Example

The error for this solution is given by

$$\| \vec{b} - A\vec{x} \| = \left\| \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} -4 \\ 3 \end{pmatrix} \right\|$$

$$= \left\| \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} - \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\|$$

$$= \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6} \approx 2.45$$

This is less than the error (=3) we calculated for our guessed best fit line earlier.

Norm is used to find the magnitude of a vector in Mathematica

In[68]:=

$$\text{Norm}\left[\begin{pmatrix}0\\0\\6\end{pmatrix}-\begin{pmatrix}1&1\\1&2\\1&3\end{pmatrix}.\begin{pmatrix}-4\\3\end{pmatrix}\right]$$

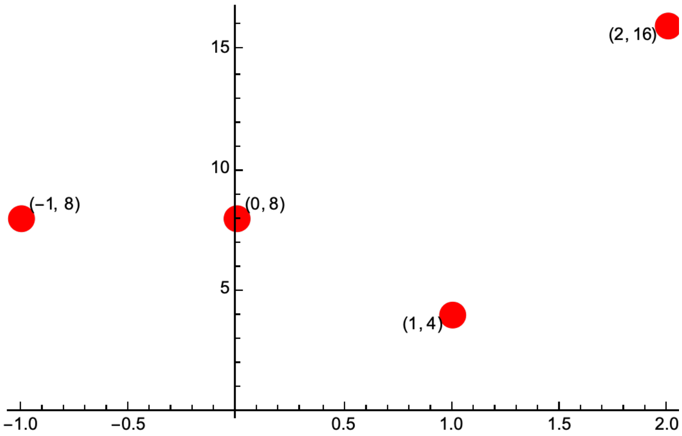
Out[68]=

$$\sqrt{6}$$

Least-square solutions for non-linear functions: Example

The least-squares approach allows you to fit other types of functions such as polynomials or exponential functions to data.

Find a **quadratic** function that best fits the following 4 data points



Example

We can express the quadratic function thus:

$f(x) = c_0 + c_1 x + c_2 x^2$

We can then use each of the 4 data points to write the following 4 equations

for data point $(-1, 8)$: $c_0 + c_1(-1) + c_2(-1)^2 = 8$

for data point $(0, 8)$: $c_0 + c_1(0) + c_2(0)^2 = 8$

for data point $(1, 4)$: $c_0 + c_1(1) + c_2(1)^2 = 4$

for data point $(2, 16)$: $c_0 + c_1(2) + c_2(2)^2 = 16$

Expressing this system of equations in matrix form

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix}$$

Example

This time (instead of using $\vec{x}$) we will let $\vec{c} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}$ giving

$A\,\vec{c} = \vec{b}$ where $A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix}$

$\vec{c}^{\,*}$ is found using the normal equation $A^T A\,\vec{c}^{\,*} = A^T\,\vec{b}$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{pmatrix}\begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 36 \\ 28 \\ 76 \end{pmatrix}$$

Example

In Mathematica we can find the RREF of the augmented matrix thus

In[69]:=
```
RowReduce[{{4, 2, 6, 36}, {2, 6, 8, 28}, {6, 8, 18, 76}}]
```

$$\text{RowReduce}\left[\begin{pmatrix} 4 & 2 & 6 & 36 \\ 2 & 6 & 8 & 28 \\ 6 & 8 & 18 & 76 \end{pmatrix}\right]$$

Out[69]//MatrixForm=

$$\begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

so $\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 3 \end{pmatrix}$

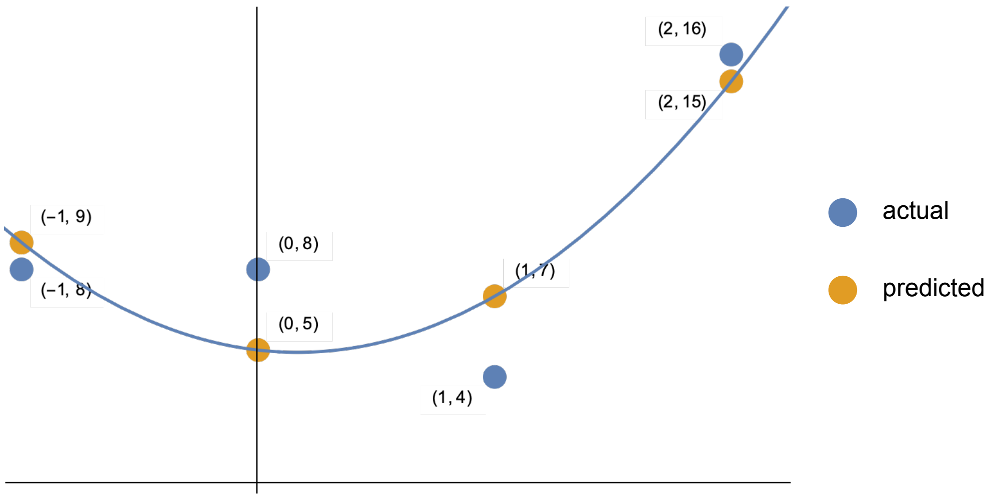
Example

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 3 \end{pmatrix}$$

The best fit quadratic to the 4 data points is

$$\boxed{f(x) = 5 - x + 3x^2}$$

Out[ ]=



We can even solve the normal equation directly in Mathematica as follows:

In[70]:=

```
A= (1 -1 1
    1  0 0
    1  1 1
    1  2 4) ; b= (8
                  8
                  4
                  16);

Solve[ (Transpose[A].A). (c0
                          c1
                          c2) == Transpose[A].b, {c0,c1,c2} ]
```

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix};\quad b = \begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix};$$

$$\text{Solve}\left[(\text{Transpose}[A].A).\begin{pmatrix} c0 \\ c1 \\ c2 \end{pmatrix} == \text{Transpose}[A].b,\ \{c0,c1,c2\}\right]$$

Out[71]//MatrixForm=

$$\begin{pmatrix} c0 \to 5 & c1 \to -1 & c2 \to 3 \end{pmatrix}$$

Terminology

Regression is a method for fitting a curve (not necessarily a straight line) through a set of points using some “best fit” criterion, typically expressed as an error function.

A **linear regression** is a regression that is linear in the unknown coefficients that are used in the fit. The most common form of linear regression is least squares fitting.

Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of an output variable based on the value of two or more other variables.

Multiple Regression Example

Consider the following observation of ice cream sales on particular days, together with the temperature and hours of sunshine on that day.

Temperature (t)	Hours Sunshine (h)	Ice Cream Sales (s)
20	6	50
25	10	100
24	2	40
15	6	10

Find a least squares fit to this data using the following prediction model

$$s(t, h) = c_0 + c_1 t + c_2 h$$

Multiple Regression Example

Temperature (t)	Hours Sunshine (h)	Ice Cream Sales (s)
20	6	50
25	10	100
24	2	40
15	6	10

Express all the data points as individual equations and represent in the form $A\vec{c} = \vec{b}$ thus

$$A = \begin{array}{c} \begin{array}{ccc} \text{always 1} & t & h \\ \downarrow & \downarrow & \downarrow \end{array} \\ \begin{pmatrix} 1 & 20 & 6 \\ 1 & 25 & 10 \\ 1 & 24 & 2 \\ 1 & 15 & 6 \end{pmatrix} \end{array}, \quad \vec{c} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}, \quad \vec{b} = \begin{array}{c} s \\ \downarrow \\ \begin{pmatrix} 50 \\ 100 \\ 40 \\ 10 \end{pmatrix} \end{array}$$

Multiple Regression Example

Form the augmented matrix of normal equation

In[72]:=
```
A=(1 20  6
   1 25 10
   1 24  2
   1 15  6);b=(50
               100
               40
               10);

(* form the augmented matrix (AᵀA⋮Aᵀb)  *)
m=Join[Transpose[A].A, Transpose[A].b, 2];

(* apply Gauss-Jordan elimination to find c0,c1,c2 *)
m=RowReduce[m]//N
```

Out[74]//MatrixForm=

$$\begin{pmatrix} 1. & 0. & 0. & -120.122 \\ 0. & 1. & 0. & 6.17886 \\ 0. & 0. & 1. & 6.72764 \end{pmatrix}$$

So $c_0 = -120.12$, $c_1 = 6.18$, and $c_2 = 6.73$

Multiple Regression Example

$c_0 = -120.122$, $c_1 = 6.17886$, and $c_2 = 6.72764$

$A\vec{c}$ gives the ice cream sales predicted by this model for our original data points

In[93]:=

$$A = \begin{pmatrix} 1 & 20 & 6 \\ 1 & 25 & 10 \\ 1 & 24 & 2 \\ 1 & 15 & 6 \end{pmatrix}; \; c = \begin{pmatrix} -120.1219512195122` \\ 6.178861788617886` \\ 6.727642276422764` \end{pmatrix};$$

```
A.c
```

Out[94]//MatrixForm=

$$\begin{pmatrix} 43.8211 \\ 101.626 \\ 41.626 \\ 12.9268 \end{pmatrix}$$

Temperature (t)	Hours Sunshine (h)	Ice Cream Sales (actual)	Ice Cream Sales (predicted)
20	6	50	43.8
25	10	100	101.6
24	2	40	41.6
15	6	10	12.9

Multiple Regression Example

Given new values of temperature (t) and sunshine (h) we can predict sales (s) thus:

$$s = \begin{pmatrix} 1 & t & h \end{pmatrix} \begin{pmatrix} -120.122 \\ 6.17886 \\ 6.72764 \end{pmatrix}$$

Consider when temperature is 20 and the hours of sunshine is 10...

In[95]:=

```
t=20; h=10;
new=(1 t h); c=(-120.122
                6.17886
                6.72764);
new.c
```

Out[97]//MatrixForm=

(70.7316)

This model predicts that the sales will be 70.7 when the temperature is 20 and the hours of sunshine is 10.

Determinants

Inverse Matrices and Determinants

Inverse matrices are widely employed in linear algebra. Remember that the solution to any matrix equation $A\vec{x} = \vec{b}$ is $\vec{x} = A^{-1}\vec{b}$, when A is invertible.

Determining when matrices are invertible is an important step in many applications. The "determinant" of a matrix is used for this purpose.

The inverse of a matrix A can be stated as:

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

Where $\det(A)$ is a special number called the determinant of A, and $\operatorname{adj}(A)$ is a special matrix called the classical adjoint matrix of A.

Ignoring how the classical adjoint matrix is found, we observe that for the inverse of A to exist its determinant cannot equal zero:

$$\det(A) \neq 0 \;\Rightarrow\; A \text{ is invertible}$$

$$\det(A) = 0 \;\Rightarrow\; A \text{ is not invertible}$$

Determinants and Eigenvalues

We will learn next week about a very important topic in linear algebra, eigenvalues and eigenvectors.

Finding the eigenvalues λ_i of a matrix A amounts to solving the following determinant equation:

$$\det(A - \lambda_i I) = 0$$

We will learn what this means next week.

The remainder of this week’s lecture concerns a range of techniques and properties to help find **determinants**.

Determinant of a 2×2 matrix

We have already seen that $A^{-1} = \frac{1}{(a\,d - b\,c)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

The determinant of the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is therefore

$$\det(A) = (a\,d - b\,c)$$

Determinant of a 1×1 matrix

It is possible to have a matrix with one row and one column.

The determinant of a 1×1 matrix A is simply the value of its single element:

$$\det(A) = a_{11}$$ when A is the 1×1 matrix (a_{11})

For example: Given $A = (5) \implies \det(A) = 5$

Determinant of a 3×3 matrix

The **determinant** of a 3×3 matrix can be found using the following formula

Given $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$

$$\det(A) = a_{11}\,a_{22}\,a_{33} + a_{12}\,a_{23}\,a_{31} + a_{13}\,a_{21}\,a_{32}$$

$$-a_{13}\,a_{22}\,a_{31} - a_{11}\,a_{23}\,a_{32} - a_{12}\,a_{21}\,a_{33}$$

Sarrus's Rule:
memory aid for 3×3 det formula

1. place the first 2 columns of the matrix A after the matrix A

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

2. add the product of the following diagonal elements

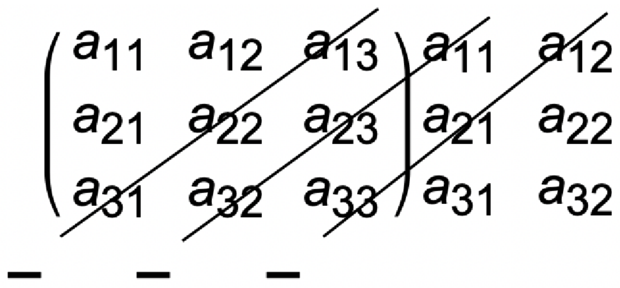
$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

$$+ \quad + \quad +$$

Giving $\longrightarrow$ $a_{11}\, a_{22}\, a_{33} + a_{12}\, a_{23}\, a_{31} + a_{13}\, a_{21}\, a_{32}$

Sarrus’s Rule contd.

3. now subtract the products of the following opposite diagonal elements



Giving ⟶

$a_{11}\, a_{22}\, a_{33} + a_{12}\, a_{23}\, a_{31} + a_{13}\, a_{21}\, a_{32}$

$-a_{13}\, a_{22}\, a_{31} - a_{11}\, a_{23}\, a_{32} - a_{12}\, a_{21}\, a_{33}$

Example

Find the determinant of $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}$

Example

$$\begin{array}{cccccc} & & & & & \end{array}\left(\begin{array}{ccc}1 & 2 & 3\\ 3 & 2 & 1\\ 1 & 2 & 1\end{array}\right)\begin{array}{cc}1 & 2\\ 3 & 2\\ 1 & 2\end{array}$$

$$- \quad - \quad - \quad + \quad + \quad +$$

$$\det(A) = 1\times2\times1 + 2\times1\times1 + 3\times3\times2 - 1\times2\times3 - 2\times1\times1 - 1\times3\times2$$

$$= 2 + 2 + 18 - 6 - 2 - 6$$

$$= 8$$

Confirmation using Mathematica

The determinants can be found using the Det function in Mathematica.

In[75]:=

```
Det[(1 2 3
     3 2 1
     1 2 1)]
```

$$\text{Det}\left[\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}\right]$$

Out[75]=

8

Notation:
$|A| = \det(A)$

For example:

$$\left|\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\right| = -2$$

$$\left|\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}\right| = 8$$

Determinants of larger $n \times n$ matrices & Laplace Expansion

Finding the determinant of larger matrices is more problematic.

The following slides introduce a method called **Laplace Expansion**, which is a **recursive method** for finding the determinant of an $n \times n$ matrix.

Cofactors

Before we can introduce this general method for calculating determinants we need to define **cofactors** ...

The i, j **cofactor** of B is the scalar $C_{i\,j}$ defined by

$$C_{i\,j} = (-1)^{i+j} \left| B_{i\,j} \right|$$

where $\left| B_{i\,j} \right|$ is the determinant of the (n – 1) × (n – 1) matrix that results from **deleting the i-th row and the j-th column** of B.

Find the cofactor C_{12} of B, where $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

Find the cofactor C_{12} of B, where B= $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

Remember:

$C_{i\,j} = (-1)^{i+j}\,\left|B_{i\,j}\right|$

where $\left|B_{i\,j}\right|$ is the determinant of the (n – 1) × (n – 1) matrix that results from deleting the i-th row and the j-th column of B.

So

$C_{12} = (-1)^{1+2}\;|B_{12}|$

$= (-1)\,\left|\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}\right|$

$= (-1)\,(4 \times 9 - 7 \times 6)$

$= 6$

The Laplace Expansion

The **Laplace expansion** is a recursive method for finding for the determinant |B| of an n × n matrix B.

Suppose $B = [b_{ij}]$ is an $n \times n$ matrix with i, j ∈ {1, 2, ..., n}.

Then its **determinant** |B| is given by:

$|B| = \sum_{j=1}^{n} b_{1j} C_{1j}$

$|B| = \sum_{j=1}^{n} b_{1j} (-1)^{1+j} |B_{1j}|$

Find the determinant $|B|$, where $B = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}$

Find the determinant $|B|$, where $B=\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}$

$|B| = \sum_{j=1}^{n} b_{1j} C_{1j}$

$= \sum_{j=1}^{n} b_{1j} (-1)^{1+j} \left|B_{ij}\right|$

$= b_{11} (-1)^{1+1} |B_{11}| + b_{12} (-1)^{1+2} |B_{12}| + b_{13} (-1)^{1+3} |B_{13}|$

$= 1 |B_{11}| - 2 |B_{12}| + 3 |B_{13}|$

$= 1 \left|\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}\right| - 2 \left|\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}\right| + 3 \left|\begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}\right|$

$= 1 (2 \times 1 - 2 \times 1) - 2 (3 \times 1 - 1 \times 1) + 3 (3 \times 2 - 1 \times 2)$

$= 1 (0) - 2 (2) + 3 (4) \qquad = 8$

Some Mathematica code that implements Laplace Expansion

This implementation has two parts:

1. A function called **ncols** that simply returns the number of columns in a matrix.
2. A recursive function **DetLaplace** that finds the determinant using Laplace Expansion.

In[76]:=
```
ncols[B_]:= Dimensions[B]⟦2⟧
```

$$\text{DetLaplace[B\_]:=}\begin{cases} \text{B⟦1,1⟧} & \text{ncols[B]==1} \\ \sum_{j=1}^{\text{ncols[B]}} \text{B⟦1,j⟧}\,(-1)^{1+j}\,\text{DetLaplace[Drop[B,\{1\},\{j\}]]} & \text{ncols[B]>1} \end{cases}$$

Mathematica implementation of Laplace Expansion

Now we can use this to find the determinant from the previous example ...

In[78]:=

$$\texttt{DetLaplace}\left[\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}\right]$$

Out[78]=

8

BUT: The time to run Laplace Expansion depends on the Factorial of the size of the Matrix!

The time to find the determinant of a randomly generated 9×9 matrix using Laplace Expansion takes several seconds on an average computer ... for a 10x10 matrix it takes 10x times longer and so on

In[83]:=

```
n=9;  A=RandomInteger[{-3,3},{n,n}];  A//MatrixForm
Print["Runtime = ", Timing[DetLaplace[A]][[1]], "secs"]
```

Out[83]//MatrixForm=

$$\begin{pmatrix} 2 & 2 & 1 & 3 & -1 & 0 & -1 & -2 & 2 \\ 3 & 3 & -2 & 0 & -1 & -1 & 1 & -3 & -3 \\ -2 & -3 & -1 & -3 & -2 & 2 & 0 & 1 & -3 \\ -3 & -1 & 0 & -2 & 0 & 0 & 0 & -1 & -3 \\ -2 & 1 & 2 & -2 & -2 & 2 & 0 & 2 & 1 \\ 2 & 0 & 1 & -3 & 0 & 0 & -2 & 1 & 2 \\ -3 & -3 & -3 & -3 & -1 & -1 & 3 & -1 & 2 \\ -2 & -2 & 1 & -2 & -3 & -1 & 1 & -2 & -2 \\ -3 & -1 & 3 & 2 & 0 & 1 & -3 & 3 & -2 \end{pmatrix}$$

```
Runtime = 2.28125secs
```

In[98]:=

```
n=10;  A=RandomInteger[{-3,3},{n,n}];  A//MatrixForm
Print["Runtime = ", Timing[DetLaplace[A]][[1]], "secs"]
```

Out[98]//MatrixForm=

$$\begin{pmatrix} 3 & 1 & 2 & 3 & -1 & 2 & 3 & 0 & 2 & 1 \\ -1 & 0 & 3 & 2 & -3 & 3 & 0 & -2 & -3 & -2 \\ 0 & 3 & 3 & -3 & 3 & 0 & -2 & 0 & -1 & 3 \\ -1 & -1 & 1 & 0 & 2 & -3 & 3 & -3 & -3 & 2 \\ -2 & -1 & -1 & -3 & -2 & -1 & -3 & 3 & -2 & -2 \\ -1 & 1 & 2 & -3 & -3 & 2 & 3 & 1 & 1 & -1 \\ -2 & 2 & -3 & -1 & -3 & -3 & 3 & -3 & 0 & 2 \\ 0 & -2 & 2 & 3 & 0 & -1 & 2 & 0 & -1 & 2 \\ 1 & 3 & 2 & -2 & -1 & -3 & -2 & 3 & -3 & 0 \\ -1 & 2 & -1 & 3 & -3 & 0 & -1 & 2 & 2 & -2 \end{pmatrix}$$

```
Runtime = 22.7651secs
```

Alternative strategies for finding the Determinants of larger $n \times n$ matrices

Laplace expansion rapidly becomes too challenging to use for pen & paper calculations

Also, due to the **factorial** nature of the method, the time to run programmed implementations rises dramatically for larger matrices, making it essentially **impractical for real-world applications**.

So, we need alternative methods to find determinants ...

Alternative strategies for finding the Determinants of larger $n \times n$ matrices

There are a range of useful strategies and rules we can use to find the determinant (quickly) for particular types of large matrices, including

- Determinants of **triangular and diagonal matrices**
- Determinants of **block matrices**
- Determinants of the **transpose**
- Determinants of **products and powers**

Determinants of triangular matrices

Triangular matrices are described as either **lower** or **upper triangular matrices**

Examples of triangular matrices

$$\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}$$

What are the determinants of the first two triangular matrices ...

What do you think the determinant of the third might be?

Determinants of diagonal matrices

Examples of diagonal matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

What are the determinants of the first two diagonal matrices ...

What do you think the determinant of the third might be?

Determinants of triangular and diagonal matrices

The determinant of an (upper or lower) **triangular matrix** or of a **diagonal matrix** is the product of the diagonal entries of the matrix.

Example

If $U = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}$ then $\det(U) = 1 \times 5 \times 8 \times 10 = 400$

confirmed in Mathematica ...

In[85]:=

$$\texttt{Det}\left[\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}\right]$$

Out[85]=

```
400
```

Example

If $D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$ then $\det(D) = 1 \times 2 \times 3 \times 4 = 24$

What is the determinant of the identity matrix I_n?

Determinant of the identity matrix

As the identity matrix I_n is a diagonal matrix with 1s for all its diagonal elements it follows that

$$\det(I_n) = 1 \times \ldots \times 1 = 1$$

Block matrices

A **block matrix** is composed of smaller matrices thus:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

Where:

- A and B have the same number of rows
- C and D have the same number of rows
- A and C have the same number of columns
- B and D have the same number of columns

Block matrices

A block matrix is composed of smaller matrices thus:

$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$

If $M = \begin{pmatrix} 1 & 2 & 1 & 2 & 3 \\ 1 & 1 & 1 & 2 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}$ and $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ what are B, C and D?

Block matrices

If $M = \begin{pmatrix} 1 & 2 & 1 & 2 & 3 \\ 1 & 1 & 1 & 2 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}$ and $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ what are B, C and D?

Answer:

$$B = \begin{pmatrix} 2 & 3 \\ 2 & 3 \\ 2 & 3 \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

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

Determinant of block matrix $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ or $\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$

If M is a **block matrix** such that $M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ or $M = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$ where A and C are **square matrices** (not necessarily of the same size) and **“0” represents a matrix of all zeros** then

$$\det(M) = \det(A)\det(C)$$

Determinant of block matrix $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ or $\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$

Find det$\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 7 & 8 & 9 & 1 & 2 \\ 10 & 11 & 12 & 3 & 4 \end{pmatrix}$

Answer

Divide the matrix into 4 blocks thus

$$\det\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 7 & 8 & 9 & 1 & 2 \\ 10 & 11 & 12 & 3 & 4 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} \det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} = 1 \times 4 \times 6 = 24$$

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ is a 2 x2 matrix}$$

$$\text{so :} \quad \det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 \times 4 - 2 \times 3 = -2$$

$$\det\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 7 & 8 & 9 & 1 & 2 \\ 10 & 11 & 12 & 3 & 4 \end{pmatrix} =$$

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} \det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 24 \times -2 = -48$$

Confirmation in Mathematica

In[86]:=

$$\text{Det}\left[\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 7 & 8 & 9 & 1 & 2 \\ 10 & 11 & 12 & 3 & 4 \end{pmatrix}\right]$$

Out[86]=

-48

Some useful determinant identities

If A is a **square matrix** then the **determinant of the transpose** $\boldsymbol{A^T}$ is given by

$$\det(A^T) = \det(A)$$

If A and B are **square matrices** then the **determinant of their product** is given by

$$\det(A.B) = \det(A).\det(B)$$