Math Camp - Day 1

Aaron Rudkin

Trinity College Dublin
Department of Political Science

rudkina@tcd.ie

Scalars and Vectors

Scalars and Vectors

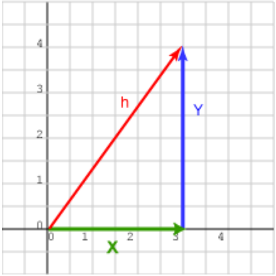
- A **scalar** is a number or element that has no inherent dimension to it. We are familiar with these:
  - 2, $\sqrt{13}$, $\frac{3}{7}$, Canada
- A **vector** combines one or more scalars, and can be interepreted geometrically as a position (i.e. having magnitude and direction):
  - $x = \{1, 4, 5\}$
  - $y = \{2, -1\}$
- Vectors are sometimes written with an arrow over their variable, as in $\vec{a}$ or $\vec{b}$
- We can refer to an individual scalar inside a vector by using subscripts:
  - $x_1 = 1, x_2 = 4, x_3 = 5$
  - $y_1 = 2, y_2 = -1$
- The **dimension** or **dimensionality** of a vector is the number of components in the vector. For x, it is 3.

Why vectors

- Suppose you are interested in studying the relationships between a series of human genes and an outcome, like a cancer diagnosis
- You recruit a number of subjects (observations) and for each, you measure various genes, as well as your outcome – this is a vector!
- Assume for now that each human gene measurement is a number: presence or absence of a gene, or number of mutations, or some numeric value
- Then we can translate those measurements into a space. More similar observations are closer together, more different observations are further apart.
- Think of regression: we want to "fit a line" through the data.

Vector Length

- Continuing with the geometric interpretation of a vector, you can think of a vector of dimensionality 2 as having an "x" and "y" coordinate in space.
- The **length** of a vector refers to how far it is from the origin in n-space, where n is the dimension of the vector.
- Length can be measured using any distance function, but the most common is **Euclidean**: $||x|| = \sqrt{\sum x_i^2}$



More on Length

- A **zero vector** is a vector whose components are all 0, thus giving the vector length 0.
- A **normalized vector** or **unit vector** is a vector whose components have been scaled so that its total length is 1. This preserves the shape and geometry of the vector, but discards the scale of the vector. Normalized vectors are written $\hat{x}$ with a "hat" over the variable.

$$\hat{x} = \frac{x}{||x||}$$

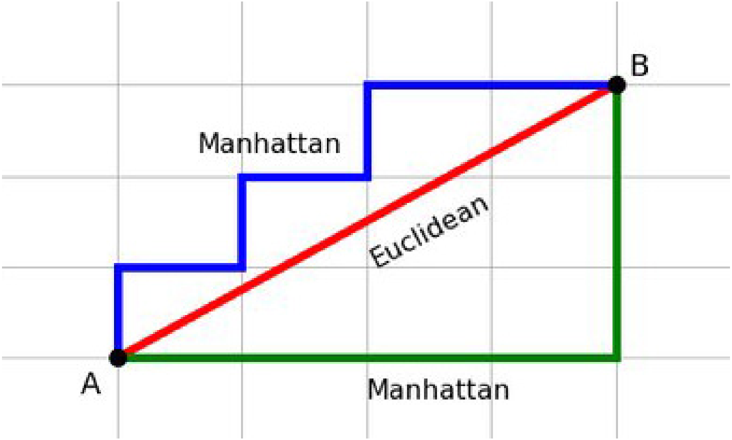
Norms of vectors

- The Euclidean length that we described above is one **norm** of a vector. Because we raise the individual magnitudes to the power of 2, we call it the ℓ_2 norm, and sometimes write it $||x||_2$.
- Another norm of interest is the **Manhattan** or **taxicab** norm (ℓ_1): The sum of absolute values of vector magnitudes: $||x||_1 = \sum_i |x_i|$.
- (Absolute value $|c|$ of a number is the positive version of the number: $|2| = |-2| = 2$

Consider $a = \{3, 4, -6, 1\}$
Then:

$$||a||_1 = 3 + 4 + |-6| + 1 = 14$$
$$||a||_2 = \sqrt{3^2 + 4^2 + -6^2 + 1^2} = \sqrt{62} = \ 7.87$$

Visualizing Manhattan versus Euclidean



Norms of vectors: p-norm

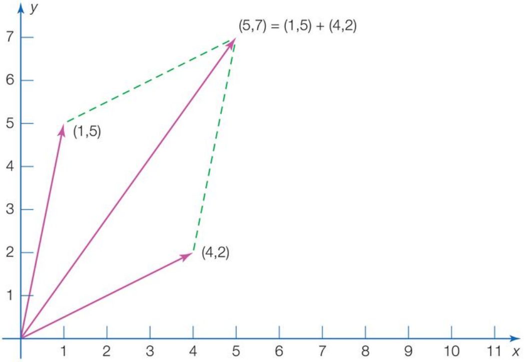
- We can generalize a **p-norm** of a vector. A p-norm of a vector is a norm, like the Euclidean norm, where we raise the values to a power, and then take the root of the subsequent values.
- $||x||_p = (\sum_i x_i^p)^{\frac{1}{p}}$
- For example, the 3-norm of $a = \{2, 4\}$ would be $||a||_3 = (2^3 + 4^3)^{\frac{1}{3}} = 4.16$
- The associated 4-norm would be $||a||_4 = (2^4 + 4^4)^{\frac{1}{4}} = 4.06$
- When just measuring length of vector, no real principled reason to choose a different norm; but later, will want distance between two vectors – choice of norm has implications for models ability to tell how similar or different observations are, especially in high dimensional space. Higher choice of norm lowers resulting number, which means bigger distances are smaller.

Vector Math

- To **add** or **subtract** two vectors, both vectors must have the same dimension. You proceed by adding or subtracting element-wise:
  - $\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, ..., a_n + b_n)$
  - $\mathbf{a}$ - $\mathbf{b} = (a_1 - b_1, a_2 - b_2, ..., a_n - b_n)$
- To add a scalar to vector, simply add it to each element: $\mathbf{a} + \mathbf{3} = (a_1 + 3, a_2 + 3, ..., a_n + 3)$
- Vectors can also be **multiplied** or **divided** by scalars:
  - $cx = (cx_1, cx_2, ..., cx_n)$
  - $\frac{x}{c} = (\frac{x_1}{c}, \frac{x_2}{c}, ..., \frac{x_n}{c})$

Vector Addition

Vector addition is associative and commutative:



Dot Product

- The most common multiplication operation involving two vectors is the **dot product** or **scalar product**
- The dot product takes in two vectors of equal dimension and outputs a scalar: a one number summary.
- This consists of the sum of the the **element-wise** products of each element of the vector:

$$a \cdot b = \sum_{i}^{n} a_i b_i$$

- $a = (1, 4, 1), b = (-2, 0, 1) \rightarrow a \cdot b = (1 \times -2) + (4 \times 0) + (1 \times 1) = -1$
- $a = (2, 4, 6), b = (3, 5) \rightarrow$ No dot product possible
- $a = (4, 6) \rightarrow a \cdot a = (4 * 4) + (6 * 6) = 52$

Note that the dot product of a vector with itself is its sum of squares

Matrices

Matrices

- If a **vector** is a series of scalars, then a **matrix** is a series of vectors. A matrix is represented as a 2D table of numbers or variables.
- If a vector is one point in some space, then a matrix is a series of points in space.
- If a vector is one observation, then a matrix is our dataset.
- Matrices are named with capital letters $\mathbf{X}$, while vectors are named with lower case letters $\mathbf{x}$.
- A matrix is made up of **rows** (horizontal) and **columns** (vertical)
- A matrix has two dimensions: $n \times m$ where n is the number of rows and m is the number of columns

How do we talk about matrices?

- A value in a matrix is subscripted X_{nm} where n is the row number from the top and m is the column from the left.

$$X_{n \times m} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

- A matrix that has one column is a **column vector**, and a matrix that has one row is a **row vector**

More Terms

- The **main diagonal** or **major diagonal** is the diagonal of the matrix heading from the top left to the bottom right
- The **minor diagonal** is the diagonal of the matrix heading from the bottom left to the top right.
- A **square** matrix is a matrix that has an equal number of columns and rows, i.e., $m = n$.
- A **zero** matrix is a square matrix in which all elements are 0.

Special Types of Matrices

- A **diagonal** matrix is a square matrix in which all elements other than those on the main diagonal are zero.
- An **identity** matrix is a diagonal matrix in which all elements on the main diagonal are 1.

$$I_{3\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- An **idempotent** matrix A is a matrix with the property $AA = A$. That is, when you multiply it by itself, it returns the original matrix – we're worry about what matrix multiplication is later!

Special Types of Matrices

- A **lower triangular** matrix has non-zero elements only on or below the main diagonal, while an **upper triangular** matrix has non-zero elements only on or above the main diagonal.

$$L_{3\times 3} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad U_{3\times 3} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$$

- A **submatrix** of a given element is the matrix that remains when we take out the row and column in which the element is (so it has one fewer column and row than the original).

Symmetric Matrices

- A **symmetric** matrix is a square matrix in which the elements are symmetric about the main diagonal, or more formally one in which $a_{ij} = a_{ji}$.

$$X_{3\times 3} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

Matrix Operations

Matrix Transposition

- The **transpose** of a matrix **X** is the the matrix where the rows and columns of **X** are switched.
- To find the transpose B of a matrix A, rewrite each element in B so that $b_{ji} = a_{ij}$.
- Notation: A^T or A'
- Dimensionality: The transpose of an $n \times m$ matrix is an $m \times n$ matrix.
- All matrices have transposes.

$$A = \begin{pmatrix} 1 & 3 & 0 \\ -1 & 6 & 2 \end{pmatrix} \qquad A^T = \begin{pmatrix} 1 & -1 \\ 3 & 6 \\ 0 & 2 \end{pmatrix}$$

Matrix Addition and Subtraction

- Two matrices can only be added to or subtracted from one another when they have the same dimensions.
- Simple: add (or subtract) the corresponding elements of the two matrices
- As a precondition, both matrices have to be of the same dimensions.
- If $A + B = C$, $c_{ij} = a_{ij} + b_{ij}$
- If $A - B = C$, $c_{ij} = a_{ij} - b_{ij}$

Transpose Properties

Inverse	$(A^T)^T = A$
Additive property	$(A+B)^T = A^T + B^T$

Additive Property Demo

Given the following matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 6 & 4 \end{pmatrix}, \; B = \begin{pmatrix} 0 & 2 \\ 1 & 5 \end{pmatrix}$$

Let's try the transposes first, then the addition:

$$A^T = \begin{pmatrix} 1 & 6 \\ 2 & 4 \end{pmatrix}, \; B^T = \begin{pmatrix} 0 & 1 \\ 2 & 5 \end{pmatrix} \rightarrow \qquad A^T + B^T = \begin{pmatrix} 1 & 7 \\ 4 & 9 \end{pmatrix}$$

Let's try the addition first, then the transpose:

$$A + B = \begin{pmatrix} 1 & 4 \\ 7 & 9 \end{pmatrix} \rightarrow \qquad (A + B)^T = \begin{pmatrix} 1 & 7 \\ 4 & 9 \end{pmatrix}$$

Scalar Multiplication

- Multiply each individual element of the matrix by the scalar to find the product.
- $C = rA$, where each $c_{ij} = ra_{ij}$.

$$3 \times \begin{pmatrix} 1 & 3 & 0 \\ -1 & 6 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 9 & 0 \\ -3 & 18 & 6 \end{pmatrix}$$

Matrix Multiplication

- To multiply two matrices, they must be **conformable**
- Conformability means the number of columns in the first matrix must match the number of rows in the second. The only matrices that can be multiplied, thus are $n \times m$ and $m \times k$ matrices.
- Can a 3×2 matrix be multiplied by a 2×3 matrix? Yes
- Can a 3×2 matrix be multiplied by a 2×4 matrix? Yes
- Can a 3×2 matrix be multiplied by a 3×3 matrix? No
- Always check for conformability before multiplying.
- The dimension of the resulting matrix will be $n \times k$, meaning it will be the number of rows from the first matrix, and the number of columns from the second.

Matrix Multiplication

- The element n, m of the result matrix will be the sum of each element in column n of the first matrix by each element of column m of the second matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ c_{41} & c_{42} & c_{43} \end{bmatrix}$$

$$c_{11} = a_{11} b_{11} + a_{12} b_{21}$$

Final notes on multiplication

- Can any matrix be multiplied by itself? No.

  Consider a 3×4 matrix.
- Can any matrix be multiplied by its transpose? Yes.

  If A is 3×4 then A' is 4×3
- Is Matrix Multiplication associative? No.

  $AB \neq BA$
- Is Matrix Multiplication associative? Yes

  $A(BC) = (AB)C$ if the matrices are conformable
- Is Matrix Multiplication distributive? Yes:

  $A(B + C) = AB + AC$ if the matrices are conformable.

Properties of Multiplication

Multiplicative property	$(AB)^T = B^T A^T$
Scalar multiplication	$(cA)^T = cA^T$

Outer Product of Vectors

- Earlier, when studying vectors, we learned the **dot product** or "inner product" $a \cdot b$, which turned the product of two vectors into a scalar.
- If we treat the vectors as a column vectors (n x 1), then the inner product can be analogized to matrix multiplication, by transposing the first vector

  $a \cdot b = A'B \rightarrow$ A transpose is a row vector
- So the resulting matrix is $(1 \times n) \times (n \times 1) = (1 \times 1)$ – a 1 by 1 matrix can be treated like a scalar.
- Another version of multiplication for vectors exists called the **outer product** $a \otimes b$, which turns the product of two vectors into a matrix.

  $a \otimes b = AB' \rightarrow$ B transpose is a row vector.
- The result will be a matrix of dimension $(n \times 1) \times (1 \times n) = (n \times n)$

Worked Example: Outer Product

$$\vec{a} = (2, 4, 1), \vec{b} = (3, 0, -1)$$

$$\vec{a} \cdot \vec{b} = 2(3) + 4(0) + 1(-1) = 5$$

$$\vec{a} \otimes \vec{b} = \begin{vmatrix} 2(3) & 2(0) & 2(-1) \\ 4(3) & 4(0) & 4(-1) \\ 1(3) & 1(0) & 1(-1) \end{vmatrix} = \begin{vmatrix} 6 & 0 & -2 \\ 12 & 0 & -4 \\ 3 & 0 & -1 \end{vmatrix}$$

Notice that the inner product is also the sum of the diagonal of the outer product!

Permutation Matrices

- A **permutation** matrix is a square matrix in which there is only a single value of 1 in any row and column, with all other elements 0.

$$P_{3\times 3} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Why is it called a permutation matrix? Because when multiplying any other matrix by a permutation matrix, the numbers are unchanged, but the rows (or columns) are scrambled:
- Given a permutation matrix P and another matrix A, the matrix PA will scramble the rows, and the matrix AP will scramble the columns

Determinants

Determinant of a Matrix

- The **determinant** of a matrix (written $|A|$) is a commonly used function that summarizes elements of a matrix as a scalar: **a determinant will just be a number**.
- Determinants are defined only for square matrices ($n \times n$).
- For $A_{2\times 2} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \rightarrow |A| = (a_{11} \times a_{22}) - (a_{12} \times a_{21})$
- In 2×2 case, main diagonal minus off diagonal.

Determinant of a Matrix

- For $n \times n$ matrices, the nightmare begins. **Laplace expansion** will be used.
- The determinant of an $n \times n$ matrix, where $n > 2$, is the sum of the products of each element and its **cofactor** for any row or column, by convention the first row

Determinant of a Matrix

- For $n \times n$ matrices, the nightmare begins. **Laplace expansion** will be used.
- The determinant of an $n \times n$ matrix, where $n > 2$, is the sum of the products of each element and its **cofactor** for any row or column, by convention the first row

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

Determinant of a Matrix

- For $n \times n$ matrices, the nightmare begins. **Laplace expansion** will be used.
- The determinant of an $n \times n$ matrix, where $n > 2$, is the sum of the products of each element and its **cofactor** for any row or column, by convention the first row

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

This is very difficult to understand, so let's take it a little bit at a time.

Determinant of a Matrix

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

- Say we have a 3x3 matrix

$$B_{3\times 3} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

Determinant of a Matrix

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

- Say we have a 3x3 matrix

$$B_{3\times 3} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

- b_{1i} refers to each of the values in the first row – b_{11}, b_{12}, b_{13}

Determinant of a Matrix

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

- Say we have a 3x3 matrix

$$B_{3\times 3} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

- b_{1i} refers to each of the values in the first row – b_{11}, b_{12}, b_{13}
- M_{1i} refers to the **minor** of the row-column. A minor is the determinant of the submatrix left when we take the row-column out.

$$M_{11} = det(\begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix}) = (b_{22} \times b_{33}) - (b_{32} \times b_{23})$$

Determinant of a Matrix

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

- Say we have a 3x3 matrix

$$B_{3\times 3} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

- b_{1i} refers to each of the values in the first row – b_{11}, b_{12}, b_{13}
- M_{1i} refers to the **minor** of the row-column. A minor is the determinant of the submatrix left when we take the row-column out.

$$M_{11} = det(\begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix}) = (b_{22} \times b_{33}) - (b_{32} \times b_{23})$$

- We multiply the minor by a sign to make it a **cofactor**: the sign of the cofactor for row i and column j is $-1^{(i+j)}$

Determinant of a Matrix

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}$$

- Say we have a 3x3 matrix

$$B_{3\times 3} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

- b_{1i} refers to each of the values in the first row – b_{11}, b_{12}, b_{13}
- M_{1i} refers to the **minor** of the row-column. A minor is the determinant of the submatrix left when we take the row-column out.

$$M_{11} = det(\begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix}) = (b_{22} \times b_{33}) - (b_{32} \times b_{23})$$

- We multiply the minor by a sign to make it a **cofactor**: the sign of the cofactor for row i and column j is $-1^{(i+j)}$

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}, \; B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i},\ B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

We will sum over $i = 1$ to 3, so:

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i},\ B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

We will sum over $i = 1$ to 3, so:

$$-1^{(1+1)} = 1,\ b_{11} = 4,\ M_{11} = det(\begin{vmatrix} 5 & 3 \\ 0 & 1 \end{vmatrix}) = (5 \times 1) - (0 \times 3) = 5$$

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}, \; B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

We will sum over $i = 1$ to 3, so:

$$-1^{(1+1)} = 1, \; b_{11} = 4, \; M_{11} = det(\begin{vmatrix} 5 & 3 \\ 0 & 1 \end{vmatrix}) = (5 \times 1) - (0 \times 3) = 5$$

$$-1^{(1+2)} = -1, \; b_{12} = -1, \; M_{12} = det(\begin{vmatrix} 4 & 3 \\ -2 & 1 \end{vmatrix}) = (4 \times 1) - (-2 \times 3) = 10$$

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}, \; B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

We will sum over $i = 1$ to 3, so:

$$-1^{(1+1)} = 1, \; b_{11} = 4, \; M_{11} = det(\begin{vmatrix} 5 & 3 \\ 0 & 1 \end{vmatrix}) = (5 \times 1) - (0 \times 3) = 5$$

$$-1^{(1+2)} = -1, \; b_{12} = -1, \; M_{12} = det(\begin{vmatrix} 4 & 3 \\ -2 & 1 \end{vmatrix}) = (4 \times 1) - (-2 \times 3) = 10$$

$$-1^{(1+3)} = 1, \; b_{13} = 1, \; M_{13} = det(\begin{vmatrix} 4 & 5 \\ -2 & 0 \end{vmatrix}) = (4 \times 0) - (-2 \times 5) = 10$$

Worked Example

$$det(B) = \sum_{i=1}^{m} (-1)^{1+i} b_{1i} M_{1i}, \; B = \begin{pmatrix} 4 & -1 & 1 \\ 4 & 5 & 3 \\ -2 & 0 & 1 \end{pmatrix}$$

We will sum over $i = 1$ to 3, so:

$$-1^{(1+1)} = 1, \; b_{11} = 4, \; M_{11} = det(\begin{vmatrix} 5 & 3 \\ 0 & 1 \end{vmatrix}) = (5 \times 1) - (0 \times 3) = 5$$

$$-1^{(1+2)} = -1, \; b_{12} = -1, \; M_{12} = det(\begin{vmatrix} 4 & 3 \\ -2 & 1 \end{vmatrix}) = (4 \times 1) - (-2 \times 3) = 10$$

$$-1^{(1+3)} = 1, \; b_{13} = 1, \; M_{13} = det(\begin{vmatrix} 4 & 5 \\ -2 & 0 \end{vmatrix}) = (4 \times 0) - (-2 \times 5) = 10$$

Putting it all together:
$det(B) = (1 \times 4 \times 5) + (-1 \times -1 \times 10) + (1 \times 1 \times 10) = 40$

Special Types of Matrices

- A **singular** matrix is one that has a determinant of 0.
- A **nonsingular** matrix has a determinant that is not 0.
- Nonsingular matrices have inverses – we'll worry about what these are tomorrow

Trace of a Matrix

- The **trace** of a square matrix is the sum of the diagonal elements:

  $\sum_i^n a_{ii}$
- Trace is more or less analogous to the derivative of the determinant, but OK to not worry about this for now

Conclusion

Tomorrow

- Matrix Inverses
- Matrices as systems of equations
- Gradients
- Hessian, Jacobian
- Matrix Calculus and OLS