ECS7024 Statistics for Artificial Intelligence and Data Science

Topic 16: Very Brief Introduction to Linear Algebra

William Marsh

Outline

- Aims
- Vectors and matrices
  - Algebra of lines
  - Shape and dimensions
- Addition
- Multiplication
  - Dot product of vectors
  - Inner and outer product
- Inverse
  - Solving linear equations
- Application: regression
- Application: multivariate normal and covariate matrix

See also the accompanying notebook

Aims

- Use of linear algebra can cause confusion
  - Very concise notation
- Many algorithms can be written using linear algebra notation
- Algorithms for linear algebra can be applied to ML

Aims is to make you aware that it is out there

Numbers

- Numbers are things we can do
  - Addition (and subtraction)
  - Multiplication (and its inverse)
- Numbers represent a quantity

- Vectors and matrices are forms of number!
  - We look at their arithmetic
  - *Problem: not obvious why it is useful*

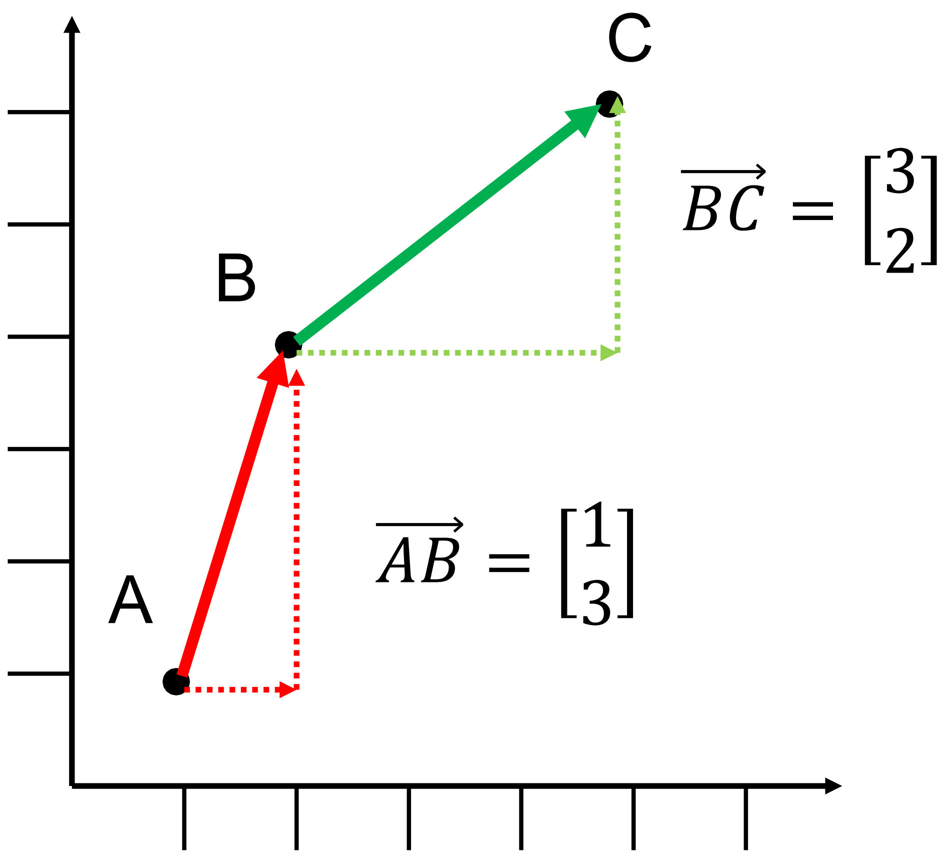
Vectors

Column Matrix

Algebra of Lines

Vectors in 2 Dimensions

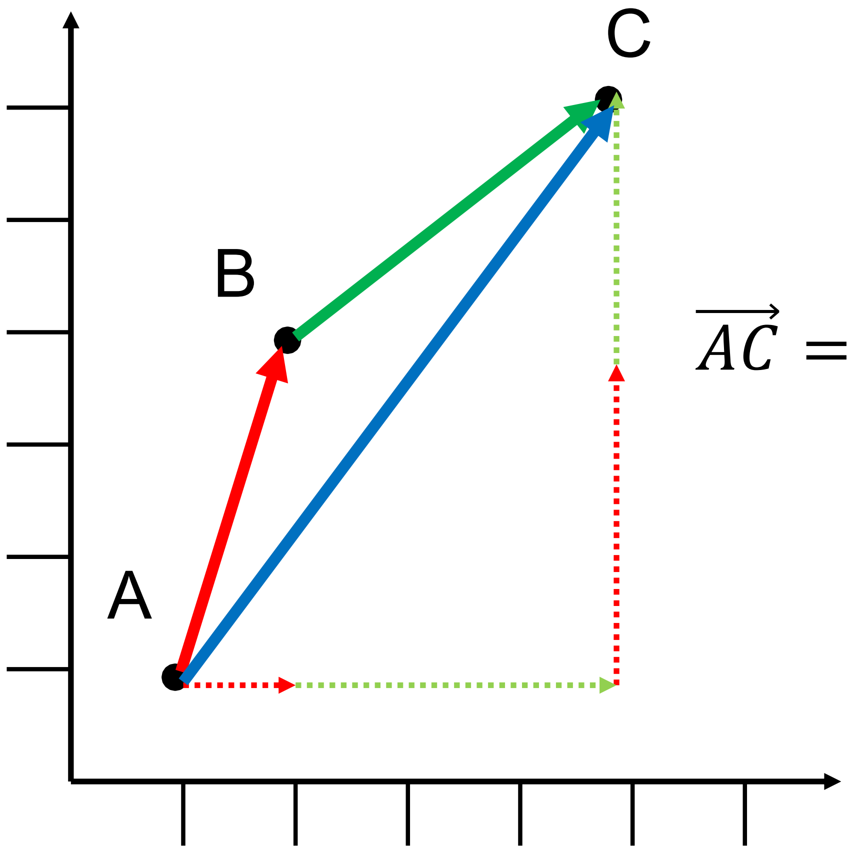
- Vectors can represent the difference between points



Vectors can be Added

- Vectors can be combined by addition

$$U = \overrightarrow{AB} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad V = \overrightarrow{BC} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



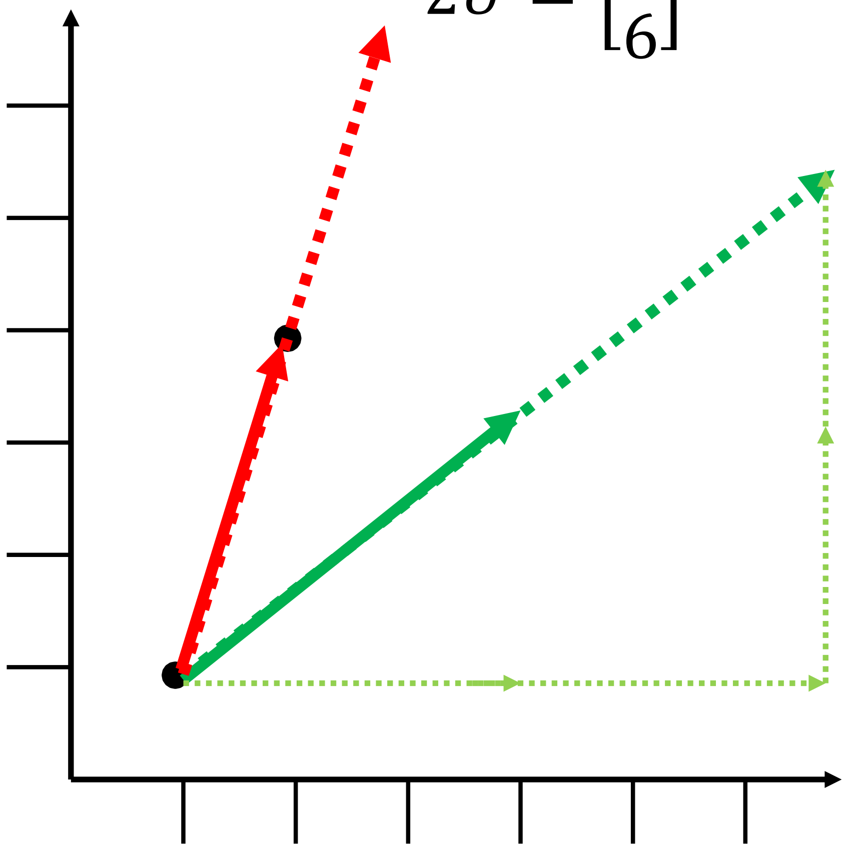
$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Multiplication By a Scalar

- Makes vector longer but does not change direction

$$2U = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$$

$$2V = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$



Length or ‘norm’ is:

$$norm(2V) = \sqrt{6^2 + 4^2}$$

Length is written as: $|V|$

Multiplication: Dot Product

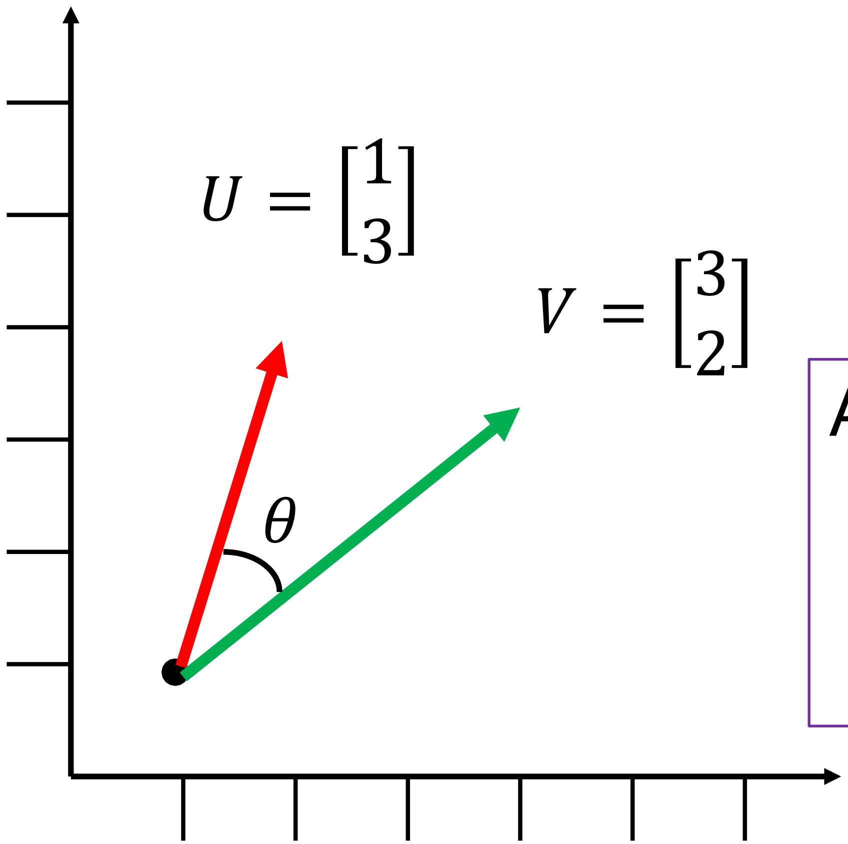
- Gives a scalar result

$$U \cdot V = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$= 1 \times 3 + 3 \times 2$$

$$U = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$V = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



Angle between vectors given by:

$$\cos(\theta) = \frac{U \cdot V}{|U||V|}$$

Bag of Words Model of Text

- Simple model of a sentence

$$\begin{array}{l} \text{word1} \\ \text{word2} \\ \text{word3} \\ \text{word4} \\ \text{word5} \\ \text{....} \end{array} \quad \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}$$

vector of counts of words

- Angle between vectors measures similarity

Matrices and Matrix Arithmetic

Shape of a Matrix

- M has
  - 3 rows
  - 2 columns
  - It is ‘3 by 2’

$$M = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix}$$

- N has
  - 2 rows
  - 3 columns
  - It is ‘2 by 3’

$$N = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Matrix Transpose

- The transpose of a matrix swaps the rows and columns
  - The transpose of M is written as M^T
  - The n^{th} column of M is the n^{th} row of M^T

$$M = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \qquad M^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix}$$

Matrix Arithmetic: Summary

- Multiply by scalar: *like a vector*
- Add
  - Provided same shape
  - *Like a vector*

- Multiplication
  - M x N is not the same as N x M
- Matrix inverse

Multiplying Matrices

- M x N is possible if
  - Columns of M equals rows of N
  - Results is ‘rows of M’ by ‘columns of N’

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1\times1+2\times4 & 1\times2+2\times5 & 1\times3+2\times6 \\ 4\times1+5\times4 & 4\times2+5\times5 & 4\times3+5\times6 \\ 7\times1+8\times4 & 7\times2+8\times5 & 7\times3+8\times6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1\times1+2\times4+3\times7 & 1\times2+2\times5+3\times8 \\ 4\times1+5\times4+6\times7 & 4\times2+5\times5+6\times8 \end{bmatrix}$$

Identity Matrix

- What is ‘1’ as in ‘1’ x M = M?
- Only for a square matrix
  - Number of rows equals number of columns

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Inverse of a Matrix

- Can we find a matrix M^{-1} so that:
  - $M \times M^{-1} = I$
  - $M^{-1} \times M = I$

- Not all matrices are invertible
- An non-invertable matrix is described as ‘singular’

Inverse of a Diagonal Matrix

- Special case: a diagonal matrix has zero values except on diagonals

$$X = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

- It is straight forward to compute the inverse (provided no diagonal is zero)

$$X^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

Conditions for Invertibility

- There are many equivalent conditions, including:
  - No two columns are same except for a scalar factor
  - No two rows are the same except for a scaler factor
  - The determinant (see below) is not zero
  - *... many more*

- In a matrix has an inverse, so does its transpose

$$(M^T)^{-1} = (M^{-1})^T$$

Simultaneous Equations using Matrices

Close connection with matrix inverse

Simultaneous Equations

- Consider the linear equations in three unknowns:

$$\begin{aligned} 4a + 2b - c &= 3 \\ -2a - b - 2c &= 0 \\ a - 5b &= 4 \end{aligned}$$

- This can be represented by:

$$\begin{bmatrix} 4 & 2 & -1 \\ -2 & -1 & -2 \\ 1 & -5 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}$$

- This is an equation of the form $\mathbf{ZX}=\mathbf{W}$
- The solution is $\mathbf{X}=\mathbf{Z}^{-1}\mathbf{W}$

Connection to Matrix Inverse

- Some simultaneous equations have no solution;
  - Some matrices have no inverse.

- Equations with 3 unknowns
  - Solution is the point of intersection
  - No solution if two of the planes are parallel (or ...)
  - Happens when any row is a linear combination of the others
- Both problems solved using Gaussian elimination

Gaussian Elimination Example

- Similar algorithm for matrix inverse

$$\begin{array}{rcl} 4a+2b-c & = & 3 \\ -2a-b-2c & = & 0 \\ a-5b & = & 4 \end{array}$$

Multiply line 2 by - ½

$$\begin{array}{rcl} 4a+2b-c & = & 3 \\ a+\frac{1}{2}b+c & = & 0 \\ a-5b & = & 4 \end{array}$$

Add lines 1 and 2

$$\begin{array}{rcl} 5a+\frac{5}{2}b & = & 3 \\ a\ -5b & = & 4 \end{array}$$

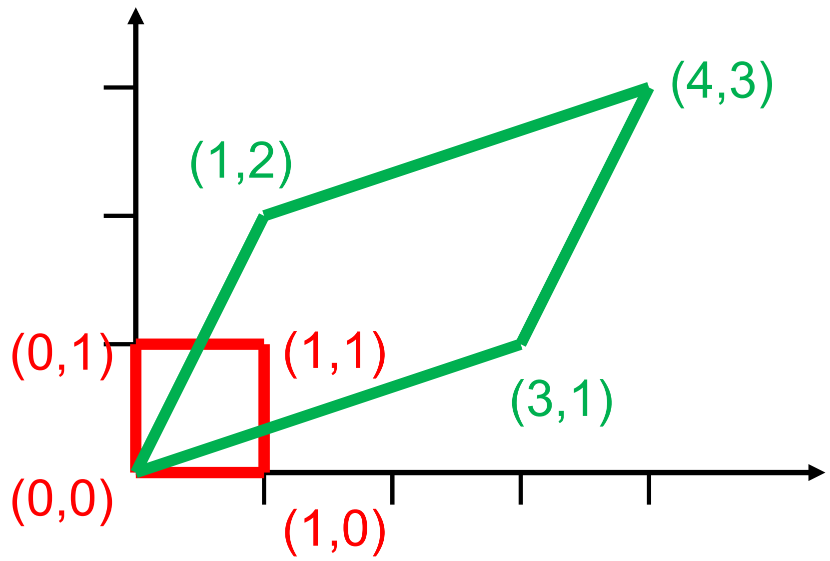
Subtract half line 2 from line 1

$$\frac{11}{2}a \quad = \quad 5$$

Matrices as Linear Transformation

Matrix as Linear Transformation

- Matrix can represent a linear transformation



represent the unit square

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 & 4 \\ 0 & 2 & 1 & 3 \end{bmatrix}$$

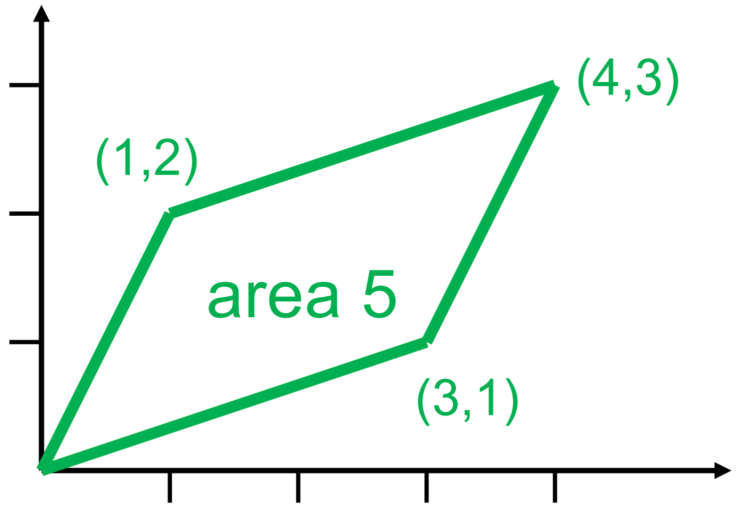
represent the transformed square

Determinant of a Matrix

- Scalar value

$$\left|\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}\right| = 3 \times 2 - 1 \times 1 = 5$$

- ... gives the area increase of the transformation



- A matrix is singular (no inverse) if the determinant is zero

Applications

Linear Regression

- Often written as:

$$y = X\beta + \epsilon$$

  - X is data matrix
  - y is result vector

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & x_{10} & x_{20} \\ 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \\ 1 & x_{14} & x_{24} \\ 1 & x_{15} & x_{25} \\ 1 & x_{16} & x_{26} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

Linear Regression II

- Applying the regression:

$$\hat{y} = X\beta$$

  - X is data matrix
  - y is result vector

$$\begin{bmatrix} \widehat{y_0} \\ \widehat{y_1} \\ \widehat{y_2} \\ \widehat{y_3} \\ \widehat{y_4} \\ \widehat{y_5} \\ \widehat{y_6} \end{bmatrix} = \begin{bmatrix} 1 & x_{10} & x_{20} \\ 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \\ 1 & x_{14} & x_{24} \\ 1 & x_{15} & x_{25} \\ 1 & x_{16} & x_{26} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Linear Regression: Finding β

- Assume there are:
  - 3 predictor variable, and 1 dependent variable
  - 4 coefficients: intercept and 3 multipliers
- We use the following variables
  - X: predictor data – 100 rows by **4** columns
  - Y: dependent data – 100 rows by 1 column
  - $\hat{\beta}$: coefficients estimated from data – 4 rows (predictors + intercept)
- The solution to the regression equation is (without justification)

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

The data transposed x data: result is 4 by 4

Inverse: also 4 by 4

... multiply by 4 by 100 giving 4 by 100

... multiply by Y: 100 by 1, giving 4 by 1

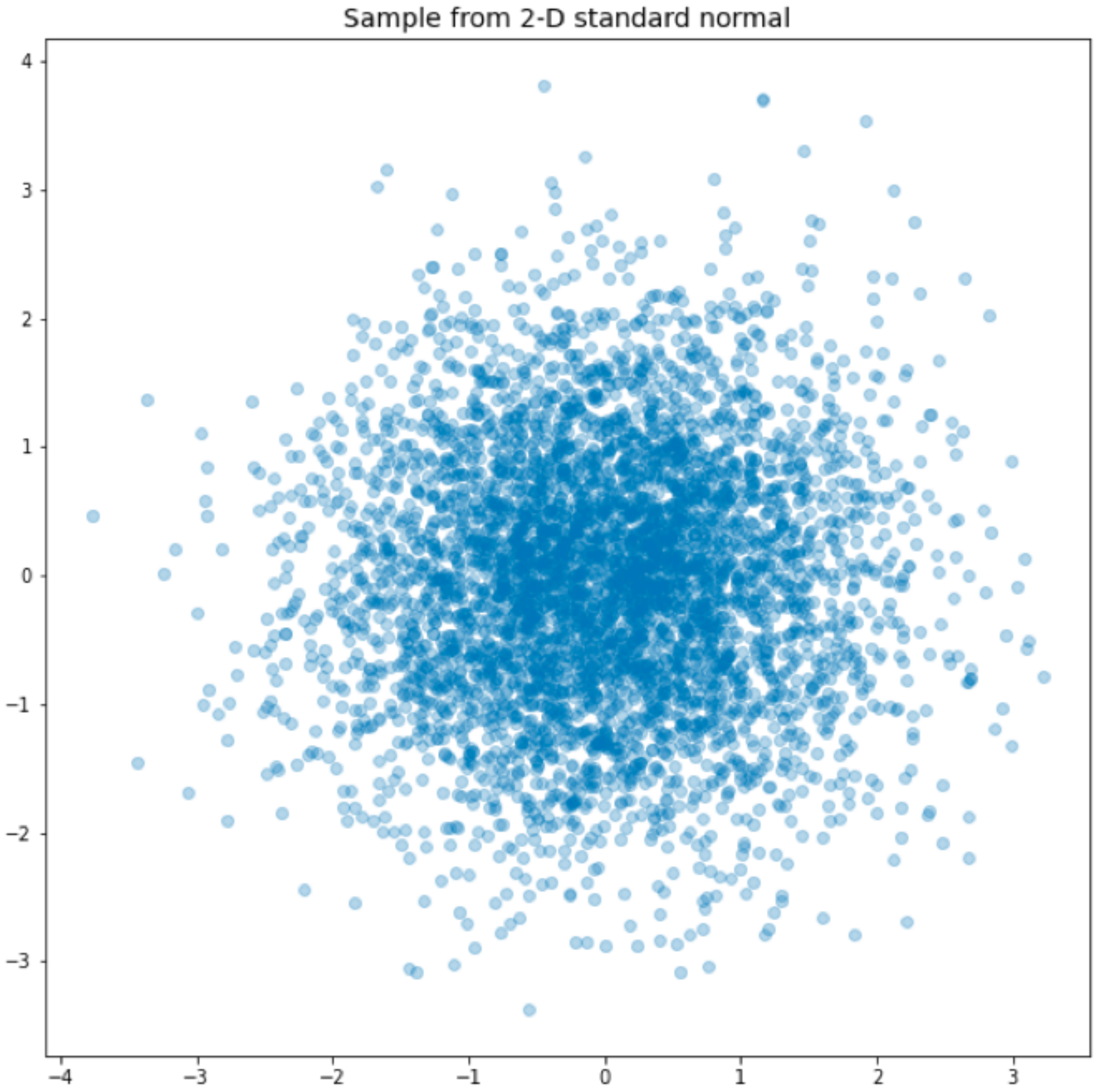
Application: Multivariate Normal

Multivariate Distribution

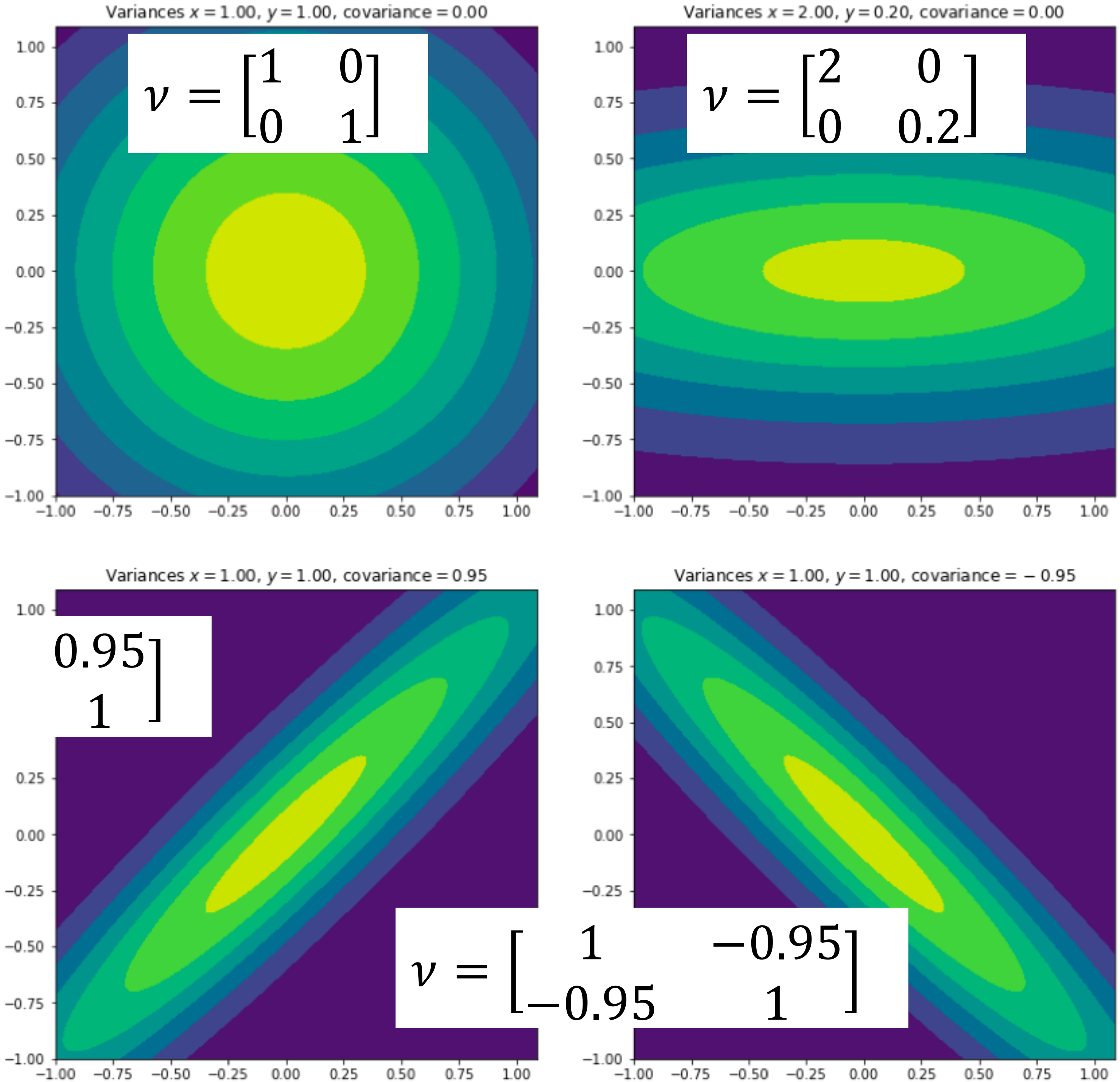
- Example: points on a target
  - Two dimensions
- Multivariate normal
  - Many dimensions
  - Each dimension has a mean – vector
  - Variance
    - Each dimension has a variance
    - Each pair of dimensions has a covariance
    - Represented by a covariance matrix

Example: Covariance Zero

- Centred on (0,0)
  - $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- Variance = 1 and covariance = 0
  - $\nu = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



$$\nu = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\nu = \begin{bmatrix} 2 & 0 \\ 0 & 0.2 \end{bmatrix}$$

$$\nu = \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1 \end{bmatrix}$$

$$\nu = \begin{bmatrix} 1 & -0.95 \\ -0.95 & 1 \end{bmatrix}$$

Summary

Summary

- Linear algebra is everywhere
  - Leads to very concise notation
  - Can be confusing
- Be aware
- Understand the arithmetic operations on matrices and vectors

Further Reading

- https://mml-book.github.io/
- Free book
- Relevant chapters
  - 2: Linear Algebra
  - 3: Analytic Geometry
  - 4: Matrix Decompositions