Linear Algebra I

Prof. Murillo
Computational Mathematics, Science and Engineering
Michigan State University

CMSE 830 Attendance Survey - Linear Algebra

Share responder link

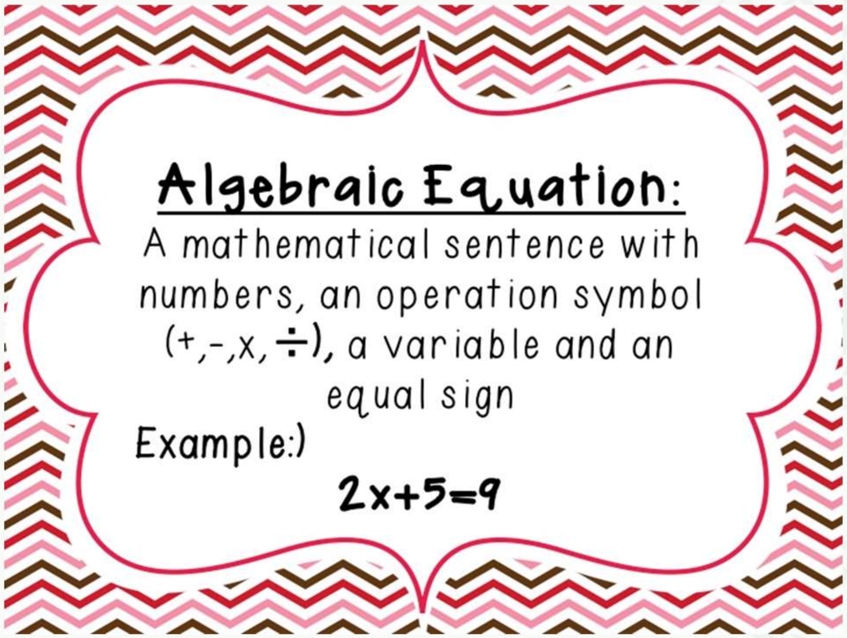
Forms for Google Docs

Download the App to create limitless forms!

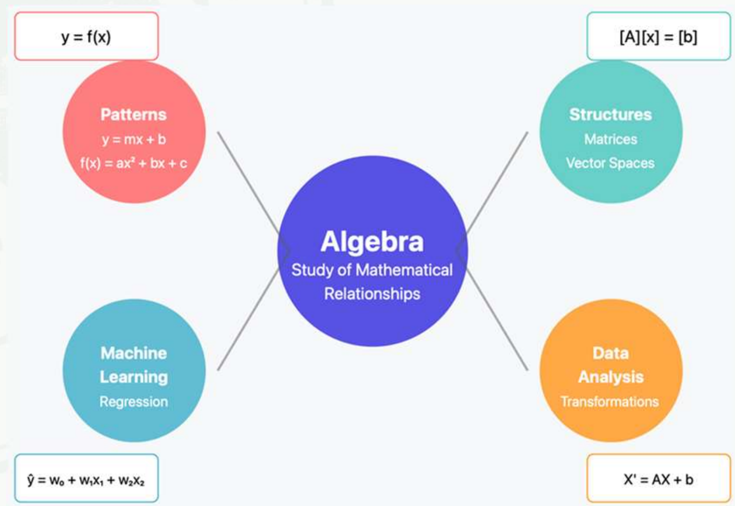
CMSE

What is an algebra?

Maybe you learned it this way?



Algebra is the branch of mathematics that studies certain abstract systems, known as algebraic structures (like matrices, vectors, and functions), and the manipulation of statements within those systems. Unlike arithmetic, which deals with specific numbers, algebra looks for general patterns and relationships that hold true regardless of the particular values involved. In data science, we use algebra to express relationships between variables, manipulate equations to solve problems, and work with structured mathematical objects like matrices that help us organize and analyze data efficiently.



Five Rules of Algebra (for Simple Numbers!)

1. Commutative Rule of Addition
2. Commutative Rule of Multiplication
3. Associative Rule of Addition
4. Associative Rule of Multiplication
5. Distributive Rule of Multiplication

$$a + b = b + a,$$

$$ab = ba,$$

$$a + (b + c) = (a + b) + c,$$

$$a(bc) = (ab)c,$$

$$a(b + c) = ab + ac$$

Not All Mathematical Objects Follow These Rules

List of algebras

From Wikipedia, the free encyclopedia

This is a list of possibly nonassociative algebras. An algebra is a module, scalars from the base ring).

- Akivis algebra
- Algebra for a monad
- Albert algebra
- Alternative algebra
- Azumaya algebra
- Banach algebra
- Birman–Wenzl algebra
- Boolean algebra
- Borcherds algebra
- Brauer algebra
- C*-algebra
- Central simple algebra
- Clifford algebra
- Cluster algebra
- Dendriform algebra
- Differential graded algebra
- Differential graded Lie algebra
- Exterior algebra
- F-algebra
- Filtered algebra

many more

For any given context, we need to establish the rules of the algebra and be sure to follow them precisely.

We need to force our thought process to fight years/decades of training.

Using software tools helps!

Five Rules of Algebra (Matrices)

1. Commutative Rule of Addition
2. Commutative Rule of Multiplication
3. Associative Rule of Addition
4. Associative Rule of Multiplication
5. Distributive Rule of Multiplication

$$a + b = b + a,$$

$$ab = ba,$$

$$a + (b + c) = (a + b) + c,$$

$$a(bc) = (ab)c,$$

$$a(b + c) = ab + ac$$

Why Linear Algebra in Data Science?

Notational and Computational Efficiency

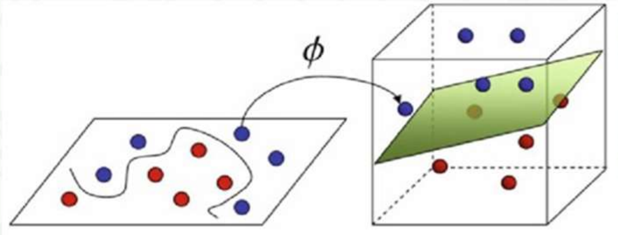
Easier to read and fewer mistakes if we avoid:

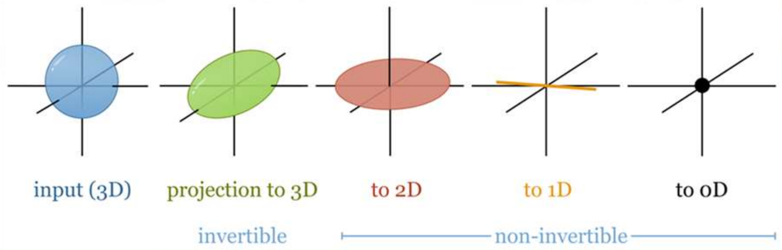
$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

New mathematical operations:

$$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathbf{y}$$

Geometric Insight





Example: Titanic Dataset

- $n = 891$ passengers
- $p = 7$ features (age, fare, class, etc.)
- each passenger is a point in $\mathbb{R}^7$
- without linear algebra: *21* separate correlation calculations
- with linear algebra: One X^TX operation

Statistics Reminder

Definitions of mean, covariance and correlation:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

average over samples of a single feature

$$\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

covariance of the feature with the “target”

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\text{Cov}(X, X)}$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Note that correlation effectively uses the *z-score*.

Definitions of mean, covariance and correlation:

$$y = w_0 + w_1 x + \epsilon$$

univariate linear model

$$w_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \text{Corr}(X, Y)\frac{\sigma_Y}{\sigma_X}$$

slope

$$w_0 = \bar{y} - w_1 \bar{x}$$

intercept/bias

The model “learns” the statistical properties of the data.

Vectors

We can organize our data into vectors.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Vectors help us write compact mathematical expressions when we have many data samples.

We can write the **transpose** as a row vector:

$$\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Inner product:

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

If we center our data (i.e., remove the mean), the covariance is written in terms of the **inner product**:

Covariance:

$$\mathrm{Cov}(X, Y) = \frac{1}{n}\mathbf{x}^T\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

Confusion Warning!

Don't confuse A^T with matrix A raised to the power T.

There are other notations for transpose, but A^T is the most common.

Idea: when you raise a matrix to a power, always use lower case.

Matrices

We almost never have a single feature. Our model would be multiple linear regression:

$$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + \epsilon$$

If we include the bias and have two features, we have N equations in three unknowns.

$$\begin{array}{cc} & \text{features} \rightarrow \\ \text{samples} \downarrow & \begin{pmatrix} x_{11} & x_{12} & \cdots \\ x_{21} & x_{22} & \cdots \\ \vdots & & \end{pmatrix} \end{array} \qquad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} \ \text{samples} \downarrow$$

vector

This introduces a **matrix**: a 2D array of numbers.

Fortunately, this is how we think about data in tidy form, as in a dataframe!

We need to define all of the basic operations among combinations of these objects.

Vector and Matrix Equations and Transpose

Vectors and matrices are very often used together to create equations:

"vector"

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

"matrix"

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}^T = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix},$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$\left[\mathbf{A}^T\right]_{ij} = \left[\mathbf{A}\right]_{ji}$$

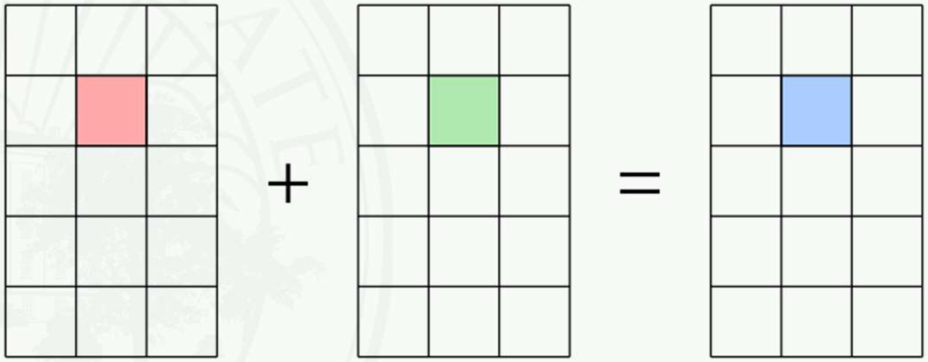
A

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

Matrix Addition

$$A + B = C$$

$$a_{ij} + b_{ij} = c_{ij}$$



Tip: Always "Call Out" Matrix Shape

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$3 \times 1 \qquad\qquad 3 \times 3 \qquad\qquad 3 \times 1$$

these are the same

Multiplication

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + a_{14}b_{41}$$

$$\begin{bmatrix} \boxed{a_{11} \quad a_{12} \quad a_{13} \quad a_{14}} \\ a_{21} \quad a_{22} \quad a_{23} \quad a_{24} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}$$

2 x 4   4 x 3   2 x 3

Back to Statistics and Commutativity

Consider a data matrix in the form:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \qquad X^T = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix}$$

We can also form the transpose X^T.

Two Covariance Matrices: X^TX versus XX^T

sum over rows

$$X^TX = \begin{bmatrix} \boxed{\sum_{i=1}^m x_{i1}x_{i1}} & \sum_{i=1}^m x_{i1}x_{i2} & \cdots & \sum_{i=1}^m x_{i1}x_{in} \\ \sum_{i=1}^m x_{i2}x_{i1} & \sum_{i=1}^m x_{i2}x_{i2} & \cdots & \sum_{i=1}^m x_{i2}x_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m x_{in}x_{i1} & \sum_{i=1}^m x_{in}x_{i2} & \cdots & \boxed{\sum_{i=1}^m x_{in}x_{in}} \end{bmatrix}$$

square nxn

sum over columns

$$XX^T = \begin{bmatrix} \boxed{\sum_{j=1}^n x_{1j}x_{1j}} & \sum_{j=1}^n x_{1j}x_{2j} & \cdots & \sum_{j=1}^n x_{1j}x_{mj} \\ \sum_{j=1}^n x_{2j}x_{1j} & \sum_{j=1}^n x_{2j}x_{2j} & \cdots & \sum_{j=1}^n x_{2j}x_{mj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n x_{mj}x_{1j} & \sum_{j=1}^n x_{mj}x_{2j} & \cdots & \boxed{\sum_{j=1}^n x_{mj}x_{mj}} \end{bmatrix}$$

square mxm

CMSE

Multiplication Of Vectors (Inner Product)

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

$$y^T z = \underbrace{\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}}_{1 \times 3} \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}}_{3 \times 1},$$

$$= y_1 z_1 + y_2 z_2 + y_3 z_3 \quad 1 \times 1$$

This operation is sometimes called a "dot product" or a "projection".

Outer Product

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} 3 \times 1,$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

$$yz^T = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \overset{1 \times 3}{\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}},$$

$$= \begin{bmatrix} y_1 z_1 & y_1 z_2 & y_1 z_3 \\ y_2 z_1 & y_2 z_2 & y_2 z_3 \\ y_3 z_1 & y_3 z_2 & y_3 z_3 \end{bmatrix} 3 \times 3$$

Note that inner products “shrink” (produce a number) and outer products expand (produce a matrix).

Multiplication by a Diagonal Matrix

$$\begin{bmatrix} -A_{11}\mathbf{r}_1- \\ -A_{22}\mathbf{r}_2- \\ \vdots \\ -A_{nn}\mathbf{r}_n- \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} -\mathbf{r}_1- \\ -\mathbf{r}_2- \\ \vdots \\ -\mathbf{r}_n- \end{bmatrix}$$

Each row of this matrix is a vector.

Sometimes written using the notation $\boldsymbol{r}^T$.

There are two important special cases:

- All of the *A*'s are 1; this is called the identity matrix *I*.
- All of the A_{mm} for $m>r$ are *0*.

Two Equations in Two Unknowns: **Determinant**

We want to solve these equations for *x* and *y*:

$$ax + by = r$$

$$cx + dy = s$$

Define the determinant:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$x = \frac{\det \begin{pmatrix} r & b \\ s & d \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}$$

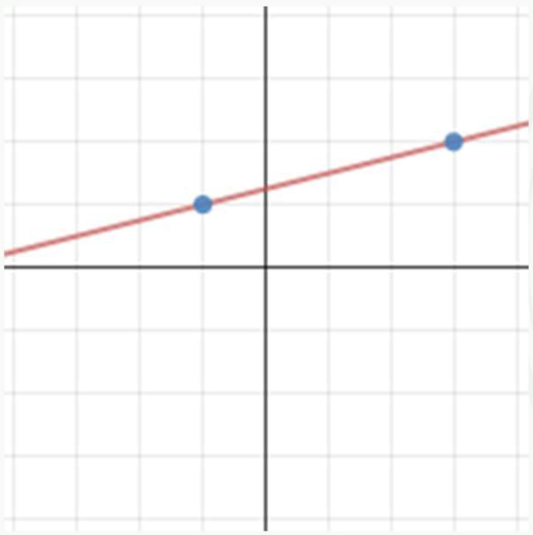
$$y = \frac{\det \begin{pmatrix} a & r \\ c & s \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}$$

Solution:

$$x = \frac{dr - bs}{ad - bc}, \quad y = \frac{as - cr}{ad - bc}$$

Interpolation: Two Equations in Two Unknowns

We want to model our data with a line. Suppose we have two points.



"bias"

$$\boxed{y = w_0 + w_1 x,}$$

model

$$y_1 = w_0 + w_1 x_1,$$

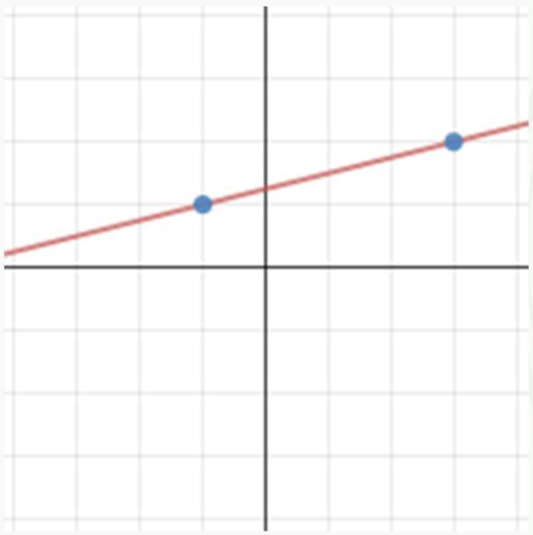
$$y_2 = w_0 + w_1 x_2,$$

Here, the weights w_0 and w_1 are our unknowns.

$$w_1 = \frac{y_2 - y_1}{x_2 - x_1} \qquad w_0 = \bar{y} - w_1 \bar{x}$$

Interpolation: Square Matrix

We want to model our data with a line. Suppose we have two points.



“bias”

$$y = w_0 + w_1 x,$$

model

$$y_1 = w_0 + w_1 x_1,$$

$$y_2 = w_0 + w_1 x_2,$$

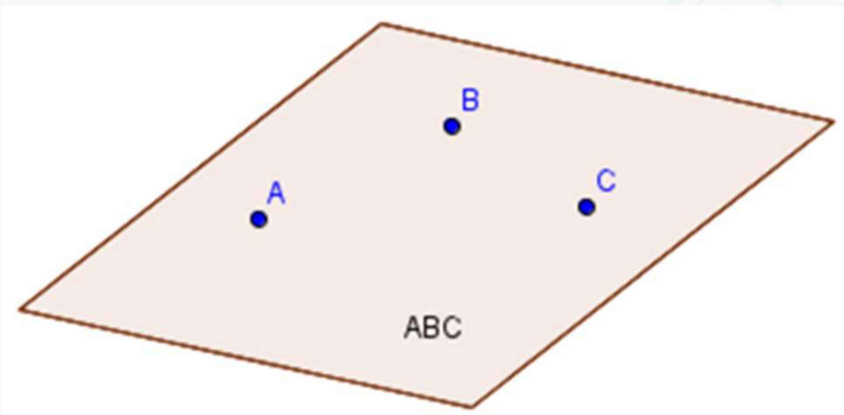
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

divide?

Multiple Linear Regression: One Bias, Two Weights

“features”

$$y = w_0 + w_1 x_1 + w_2 x_2,$$



$$y_1 = w_0 + w_1 x_{11} + w_2 x_{21},$$

$$y_2 = w_0 + w_1 x_{21} + w_2 x_{22},$$

$$y_3 = w_0 + w_1 x_{31} + w_2 x_{32},$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ 1 & x_{13} & x_{23} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

In data science, this matrix will almost never be square.

Division? That's what we really need!

we want to solve for **w** $\mathbf{y} = K\mathbf{w},$

$K\mathbf{w} = \mathbf{y},$ just rewrite the equation

multiply both sides by *M* $MK\mathbf{w} = M\mathbf{y},$

$MK = I,$ choose M such that this is true

I is the "identity matrix" $$I = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & & & \end{bmatrix}$$

$K^{-1} \equiv M,$ call *M* the inverse of *K*

we have our solution! $\mathbf{w} = K^{-1}\mathbf{y}$

The challenge is finding the inverse!

$$K^{-1}K = I$$

ICA Example 1: Model is a Line

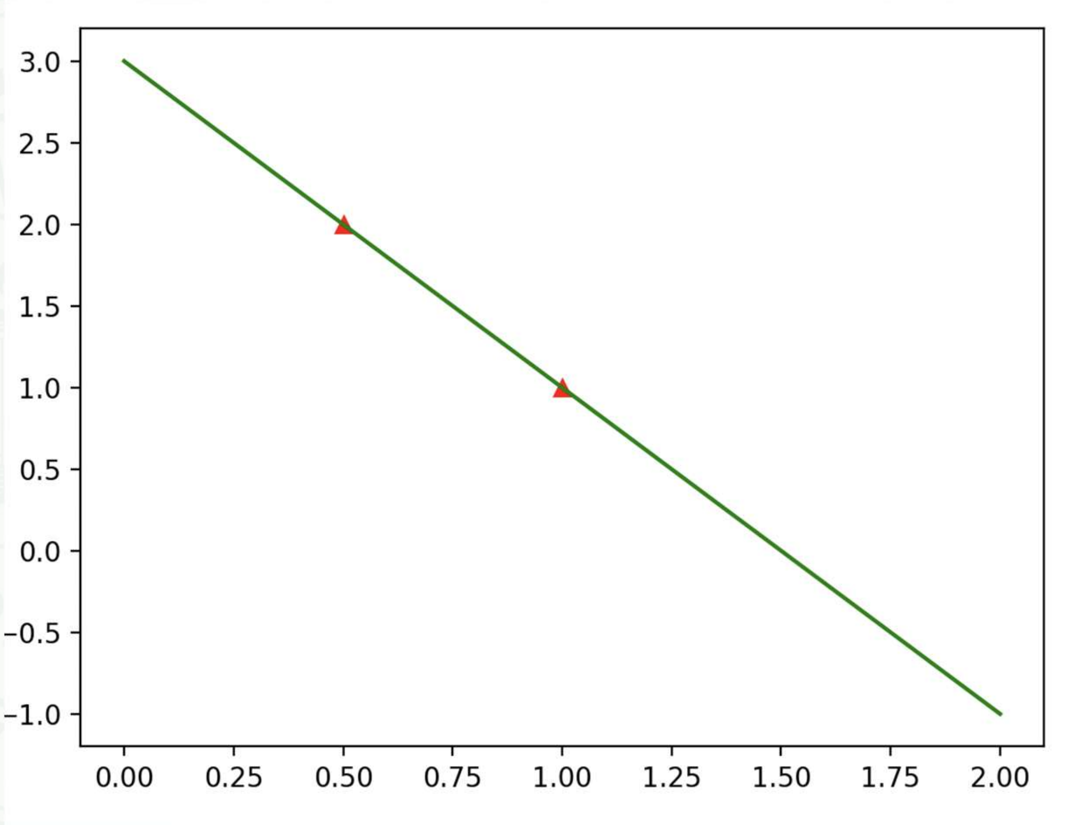
Our two data points:

(0.5, 2), (1, 1)

$$2 = w_0 + w_1 0.5$$

$$1 = w_0 + w_1 1$$

$$\boxed{y = 3 - 2x}$$



ICA Example 2: Model is an RBF-NN

Our two data points:

(0.5, 2), (1, 1)

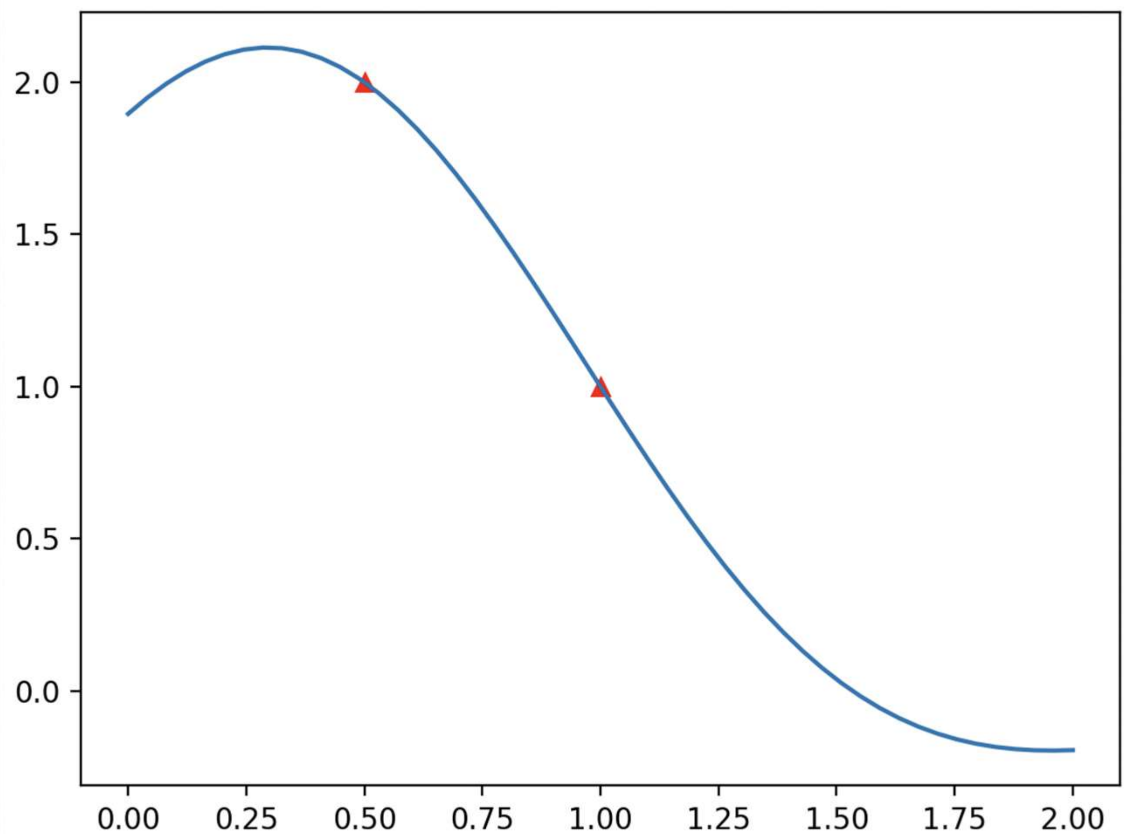
$$\boxed{y = w_0 e^{-(x-x_1)^2} + w_1 e^{-(x-x_2)^2}}$$

$$2 = w_0 + w_1 e^{-(-0.5)^2}$$

$$1 = w_0 e^{-(0.5)^2} + w_1$$

$$w_0 = \frac{2-A}{1-A^2}$$

$$w_1 = \frac{2-w_0}{A}$$

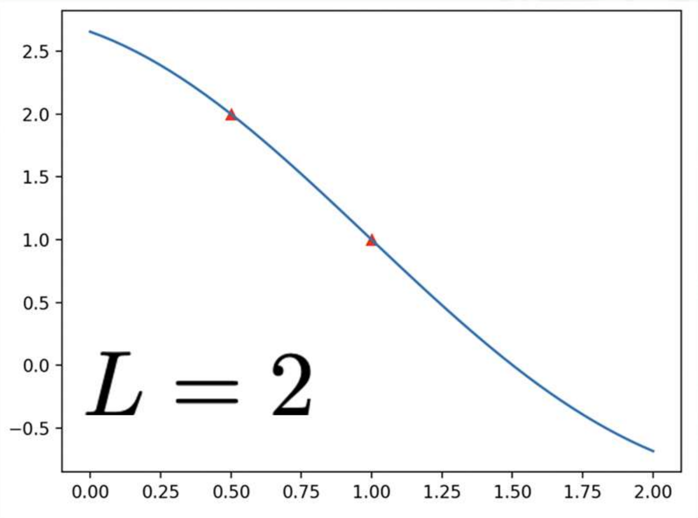


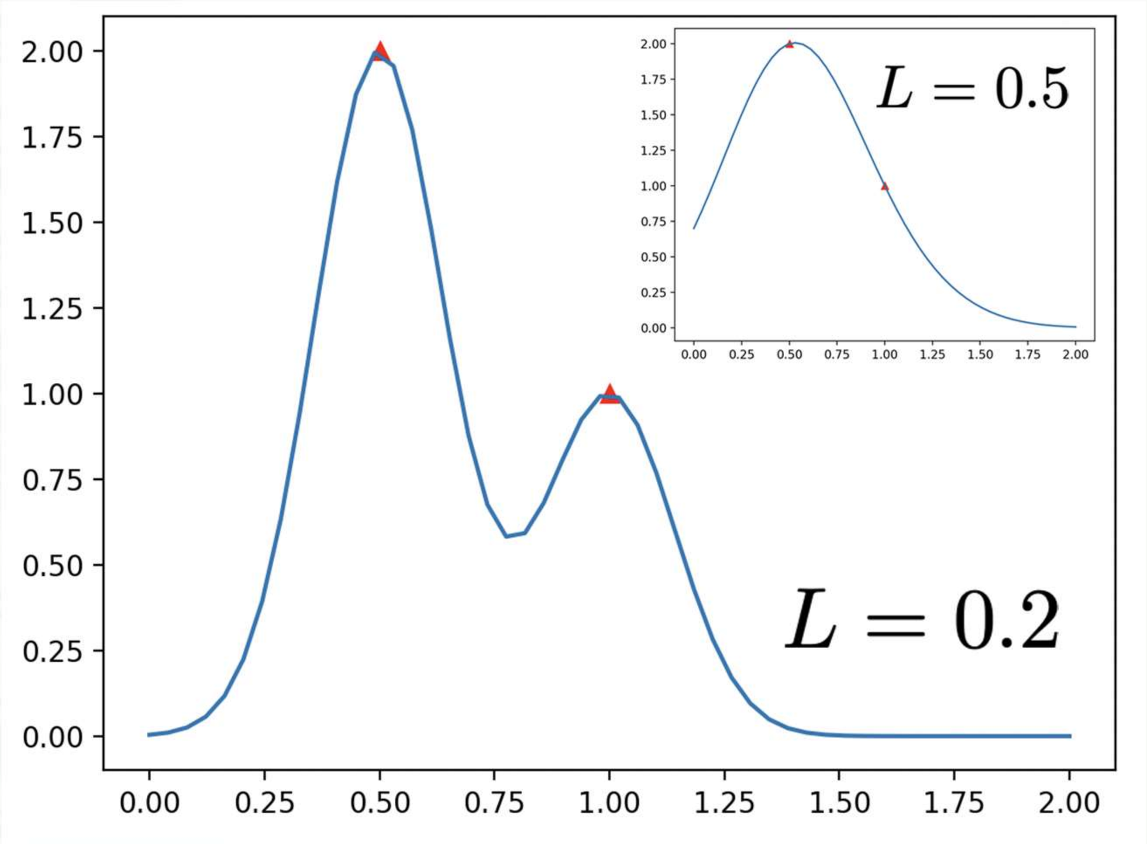
ICA Example 2: Model is an RBF-NN

Our two data points:

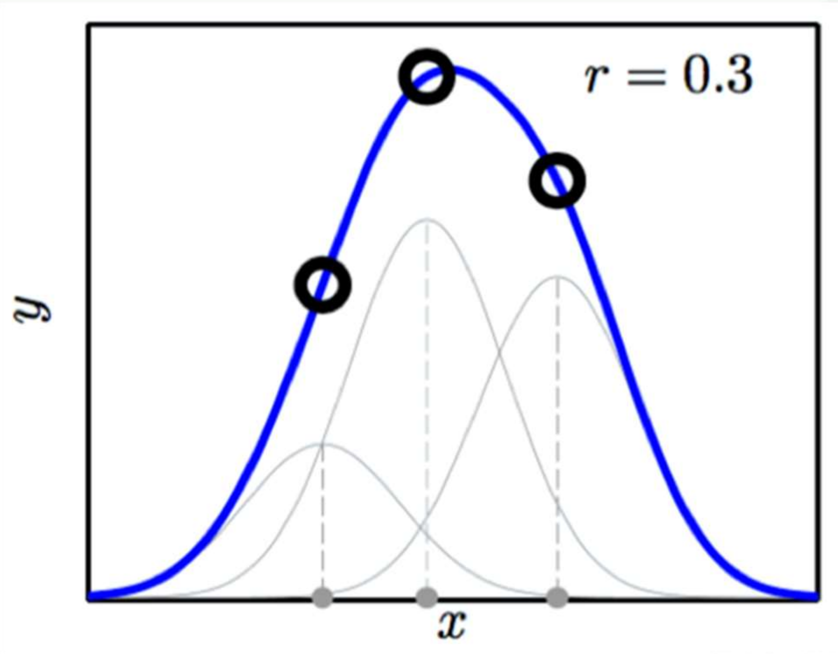
(0.5, 2), (1, 1)

$$y = w_0 e^{-(x-x_1)^2/L^2} + w_1 e^{-(x-x_2)^2/L^2}$$





Radial Basis Function Neural Networks



$$y(x) = w_1 K(x, x_1) + w_2 K(x, x_2) + w_3 K(x, x_3),$$
$$y(x_1) = w_1 K(x_1, x_1) + w_2 K(x_1, x_2) + w_3 K(x_1, x_3),$$
$$y(x_2) = w_1 K(x_2, x_1) + w_2 K(x_2, x_2) + w_3 K(x_2, x_3),$$
$$y(x_3) = w_1 K(x_3, x_1) + w_2 K(x_3, x_2) + w_3 K(x_2, x_3),$$
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

The math problem is still the same!

Pick whatever *K* makes sense for your modeling!

Non-Square (Realistic!) Matrices

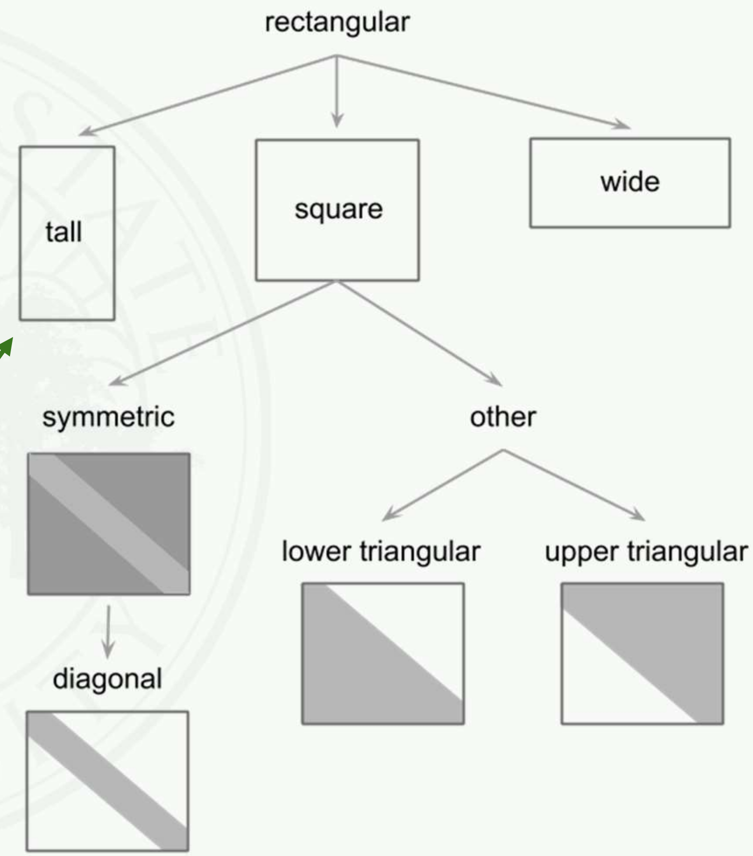
Next week, and in the ICA, we use a more general approach to the inverse.

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

You will implement this using Numpy's `linalg` library.

```
   w = np.linalg.inv(X.T @ X) @ X.T @ y
or w = np.linalg.pinv(X) @ y
```

Most common in data science.



Quadratic Form

$$(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = 0$$

For now, we set the quadratic form to 0.

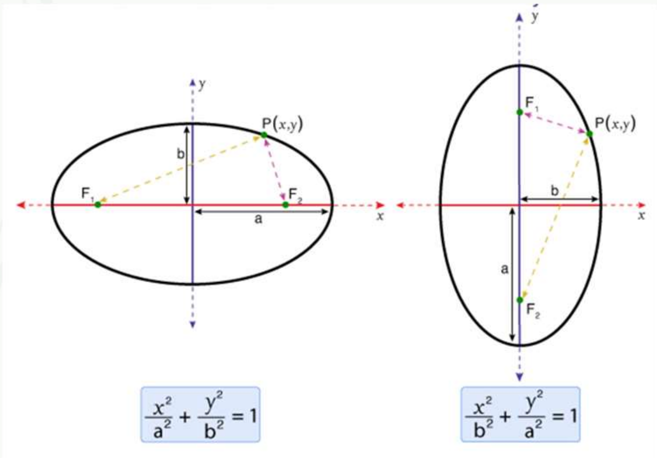
$$(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = 0$$

$$\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} = 0$$

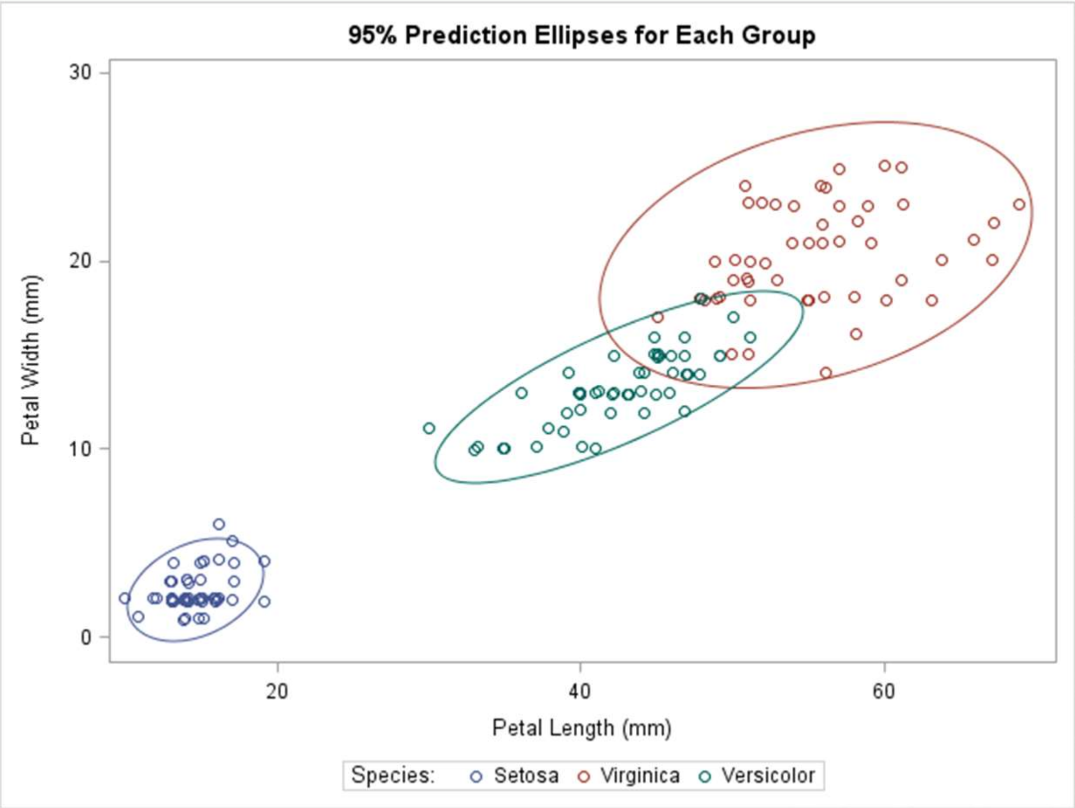
Assume diagonal A^{-1}.

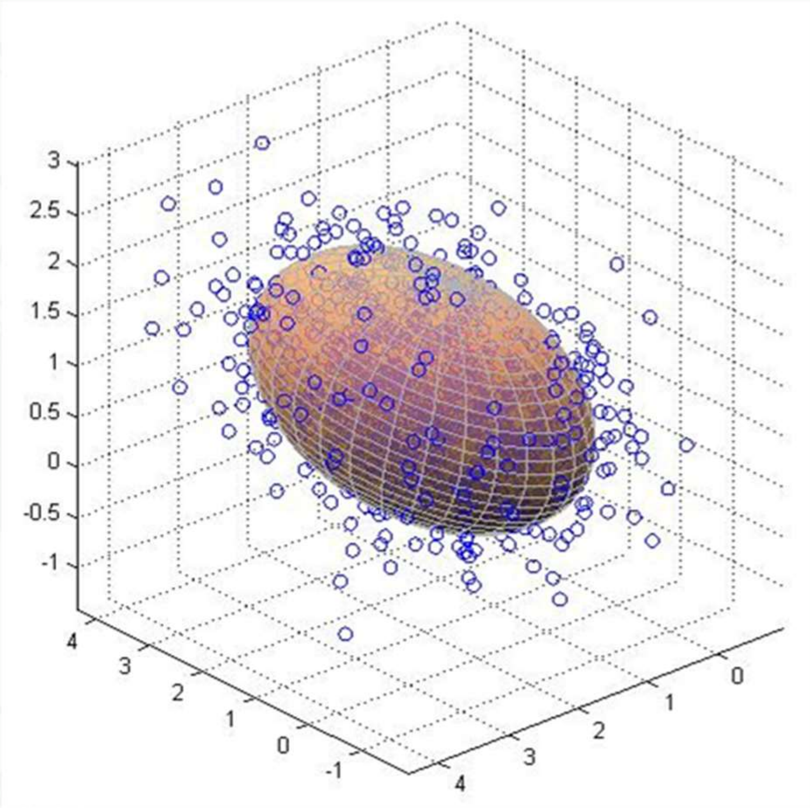
$$\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} = 0$$

$$a_{11}(x_1-\mu_1)^2 + a_{22}(x_2-\mu_2)^2 = 0$$



You Can Build Ellipsoids in Any Dimensions >1





In general, *A* is **not** diagonal and the ellipsoid is rotated.

Building Multivariate Gaussians With Matrices

$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

Mahalanobis Distance