Notes based off of Zico Kolter's notes on Linear Algebra

Concepts & Notation

Consider the following system of equations:

$$\left.\begin{aligned} 4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9 \end{aligned}\right\} \begin{array}{l} 2 \text{ equations} \\ 2 \text{ variables} \end{array}$$

A is a mapping.

Discuss column view for $Ax=b$

$\Downarrow$ more compactly

$$Ax = b$$

$$\begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

Solving for A

$$A^{-1}Ax = A^{-1}b$$

$$Ix = A^{-1}b$$

$$x = \begin{bmatrix} -3.8182 \\ .4545 \end{bmatrix}$$

Notation:

$A \in \mathbb{R}^{m \times n}$ — m: rows, n: columns

$x \in \mathbb{R}^n$ — n.rows

↳ column vector

x^T

↳ row vector

x_i = i^{th} element

A_{ij} refers to the i^{th} row and the j^{th} column

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix}$$

a_1

$$\begin{bmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \cdots & a_n \\ | & | & | & & | \end{bmatrix}$$

$A_{:j}$ or a_j as the j^{th} column of A.

$A_{i:}$ or a_i^T

$$A = \begin{bmatrix} — a_1^T — \\ — a_2^T — \\ \vdots \\ — a_m^T — \end{bmatrix}$$

Note: $a_1^T \neq a_1$

ambiguous definitions

Vector - Vector Products

Given 2 vectors $x, y \in \mathbb{R}^n$, the quantity $x^T y$ is sometimes called the inner product or dot product.

$$x^T y \in \mathbb{R} = [x_1, x_2, \dots x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Note that $x^T y = y^T x$

$$\underset{N \times 1}{x} \quad \underset{1 \times N}{x^T} \quad \underset{N \times 1}{y}$$

Outer product

$x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ then $xy^T \in \mathbb{R}^{m \times n}$ is called the outer product

$$\underset{n \times 1 \;\; 1 \times n}{xy^T} \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} [y_1 \; y_2 \; \dots \; y_n] = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \vdots & & & \vdots \\ x_m y_1 & x_m y_2 & \dots & x_n y_m \end{bmatrix}$$

Compact representation

$1 \in \mathbb{R}^n$

$x \in \mathbb{R}^m$

$$A = \begin{bmatrix} x_1 & x_1 & \dots & x_1 \\ \vdots & \vdots & & \vdots \\ x_n & x_n & & x_m \end{bmatrix} \Rightarrow \underset{m \times 1}{x} \, \underset{1 \times N}{1^T}$$

$m \times n$ matrix

Matrix Vector products

matrix $A \in \mathbb{R}^{m \times n}$

$x \in \mathbb{R}^n$

$y = Ax \in \mathbb{R}^m$

$$y = Ax = \begin{bmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_m^T - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ a_3^T x \\ a_m^T x \end{bmatrix}$$

In other words, the ith entry of y is equal to the inner product of the ith row of A and x.

Alternative representation

Linear combo columns of A

m entries

$$y = \overset{m\times n}{A}\,\underset{n\times 1}{x} = \begin{bmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \cdots & a_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \underbrace{\underbrace{\begin{bmatrix} \\ a_1 \\ \\ \end{bmatrix}}_{m\times 1} \underbrace{x_1}_{1\times 1}}_{m\times 1} + \begin{bmatrix} \\ a_2 \\ \\ \end{bmatrix} x_2 + \cdots \begin{bmatrix} \\ a_n \\ \\ \end{bmatrix} x_n$$

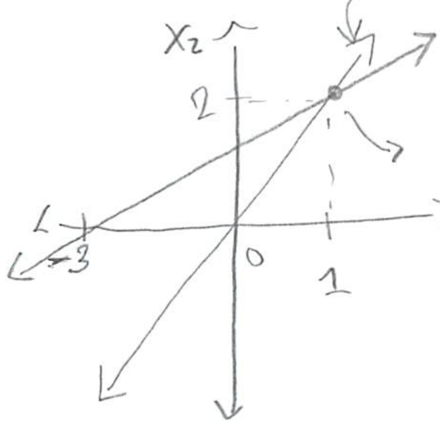
Some interesting matrix properties

1) Matrix multiplication is associative $(AB)C = A(BC)$
2) Matrix multiplication is distributive $A(B+C) = AB + AC$
3) Matrix multiplication (in general) is not commutative
$AB \neq BA$

Row Picture

$2x_1 - x_2 = 0$
$-x_1 + 2x_2 = 3$

$x_1 = 1$
$x_2 = 2$



Here $\quad 2\times 2 \quad 2\times 1 \Rightarrow 2\times 1$

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

varying this from $[-\infty, \infty]$ will fill up the entire plane

$$\Rightarrow \begin{bmatrix} 2 \\ -1 \end{bmatrix} x_1 + \begin{bmatrix} -1 \\ 2 \end{bmatrix} x_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

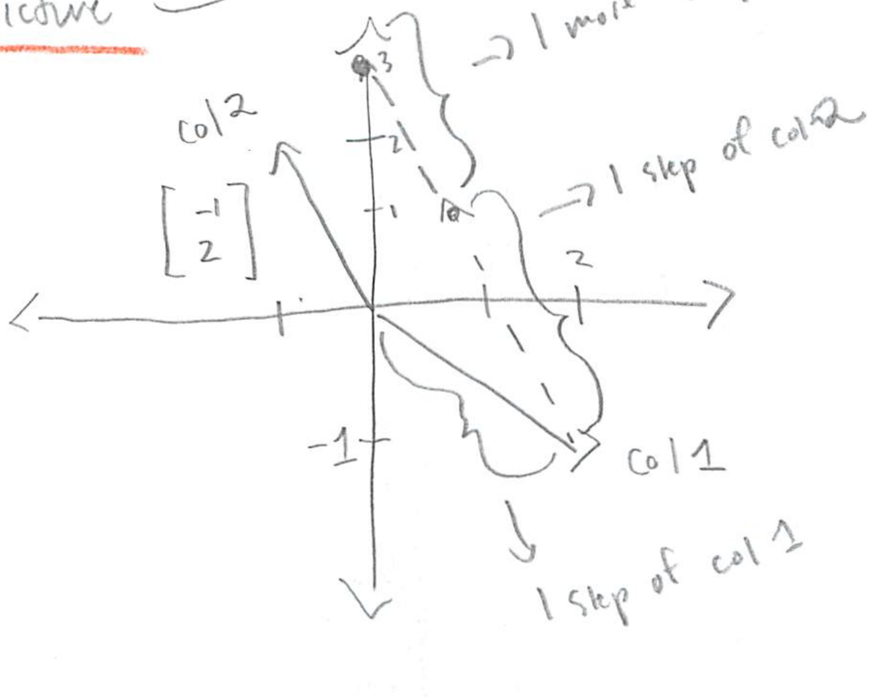
go down 2

go down 1

go left 1

go up 2

Column Picture



The identity matrix and diagonal matrices

identity matrix denoted $I \in \mathbb{R}^{n \times n}$ is a square matrix with ones on the diagonal.

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

For all $A \in \mathbb{R}^{m \times n}$

$$\hookrightarrow \underbrace{A \overset{n \times n}{I} = A = \overset{m \times m}{I} A}_{\uparrow \text{ Is this true?}}$$

Question

What does a multivariate Gaussian with an identity matrix represent?

Diagonal Matrices

all non-diagonal elements are 0

$$D = \operatorname{diag}(d_1, d_2, \dots, d_n)$$

with $D_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$

The transpose of a matrix

Given a matrix $A \in \mathbb{R}^{n \times m}$ its transpose is $A^T \in \mathbb{R}^{m \times n}$

$$(A^T)_{ij} = A_{ji}$$

Properties of the transpose:

1) $(A^T)^T = A$

2) $(AB)^T = B^T A^T$

3) $(A + B)^T = A^T + B^T$

Symmetric Matrices

$A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$.

It is anti-symmetric if $A = -A^T$.

The matrix $A + A^T$ is symmetric? Ask Question

Trace

The trace of a square matrix $A \in R^{n \times n}$ denoted $tr(A)$ is the sum of the diagonal elements in the matrix.

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

For $A \in R^{n \times n}$, $tr(A) = tr(A^T)$

For $A, B \in R^{n \times n}$, $tr(A+B) = tr(A) + tr(B)$

For $A \in R^{n \times n}$, $s \in R$, $tr(sA) = s \times tr(A)$

For A, B s.t $\underset{4\times7 \; 7\times4}{AB}$ is square, $tr(\underset{4\times7}{A}\underset{7\times4}{B}) = tr(\underset{7\times4}{B}\underset{4\times7}{A})$

For A, B, C s.t ABC is square, $tr(\underset{3\times5}{A}\underset{5\times6}{B}\underset{6\times3}{C}) = tr(BCA) = tr(CAB)$

Norm

The norm of a vector $\|x\|$ is informally the "length" of the vector. We have the euclidean or L2 norm.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

More formally, a norm is any function $f: R^n \to R$ that satisfies 4 properties:

1) For all $x \in R^n$, $f(x) \geq 0$ (non-negativity)
2) $f(x) = 0$ iff $x = 0$ (definiteness)
3) For all $x \in R^n$, $t \in R$, $f(tx) = |t| f(x)$ (homogeneity)
4) For all $x, y \in R^n$, $f(x+y) \leq f(x) + f(y)$ (triangle inequality)

L_1 Norm $\Rightarrow \|x_1\| = \sum_{i=1}^{n} |x_i|$

Family of l_p norms, parameterized by a real number $p \geq 1$ and defined as:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Norms also exist for matrices, such as the Frobenius norm.

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{tr(A^T A)}$$

Linear Independence and Rank

A set of vectors $\{x_1, x_2, \dots x_n\} \subset R^m$ is said to be linearly independent if no vector can be represented as a linear combination of the remaining vectors.

Linearly dependent if:

some scalar value

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

$$\alpha_i \in R$$

example:

$x_3 = -2x_1 + x_2$

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Column rank of $A \in R^{m \times n}$ is the largest subset of columns of A that constitute a linearly independent set.

equal to the row rank. Just call it the "rank".

Properties of the "rank"

1) For $A \in R^{m \times n}$, $\text{rank}(A) \leq \min(m,n)$.

If $\text{rank}(A) = \min(m,n)$, then A is full rank.

2) For $A \in R^{m \times n}$, $\text{rank}(A) = \text{rank}(A^T)$

3) For $A \in R^{m \times n}$ and $B \in R^{n \times p}$, $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$

4) For $A, B \in R^{m \times n}$, $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$

Inverse

The inverse of a square matrix $A \in R^{n \times n}$ is denoted A^{-1} and represents a unique matrix s.t:

$$A^{-1}A = I = AA^{-1}$$

If A has an inverse it is called a non-singular matrix or an invertible matrix.

For A to be non-singular, then A must be full rank!!

Properties of the inverse A^{-1}, B^{-1} (A and B are non singular) $\in R^{n \times n}$

1) $(A^{-1})^{-1} = A$

2) $(AB)^{-1} = B^{-1}A^{-1}$

3) $(A^{-1})^T = (A^T)^{-1} = A^{-T}$

4) $(kA)^{-1} = k^{-1}A^{-1}$

When is a matrix singular?

When $|A| = 0$

the determinant

Inverse for a 2×2 $\Rightarrow$ $A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \dfrac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$$= \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Orthogonality

Two vectors $x, y \in R^n$ are orthogonal if $\underbrace{x^T y}_{\text{inner product}} = 0$

A square matrix $U \in R^{n \times n}$ is orthogonal if all its columns are orthogonal to each other and are normalized. orthonormal columns

↳ if $\|x\|_2 = 1$, then it is normalized

1) $U^T U = I = U U^T \Rightarrow$ this means that the inverse of an orthogonal matrix is its transpose.

· Operating on an orthogonal matrix will not change its norm $\Rightarrow \|Ux\|_2 = \|x\|_2$

Range & Nullspace of a Matrix

The span of a set of vectors $\{x_1, x_2, \ldots x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, \ldots x_n\}$.

$$\text{span}(\{x_1, \ldots x_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i x_i, \ \alpha_i \in R \right\}$$

If $x_1 \ldots x_n$ are linearly independent, then the span of this set encompasses the whole set of real numbers R^n. Any vector $v \in R^n$ can be written as a linear combination of x_1 through x_n.

The set of all possible combinations is called the column space or the range of a matrix A with column vectors $x_1 \ldots x_n$.

Nullspace of a matrix $A \in R^{m \times n}$, denoted $N(A)$ is the set of all vectors that equal 0 when multiplied by A.

$$N(A) = \{x \in R^n : Ax = 0\}$$

The determinant

A function $R^{n \times n} \to R$ for some matrix $A \in R^{n \times n}$ denoted as $|A|$ or $\det A$.

Consider this matrix A

$$\begin{bmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_n^T - \end{bmatrix}$$

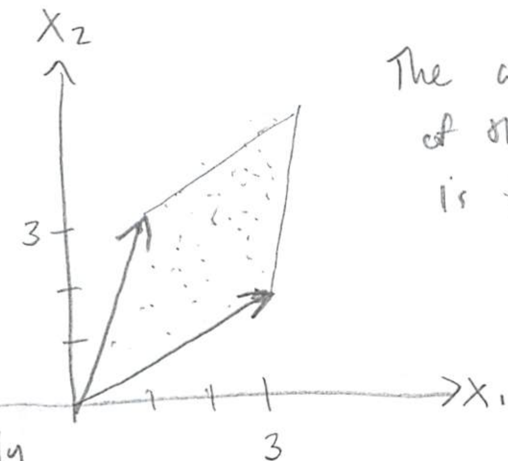
Consider the set of points $S \subset R^n$ formed by taking all possible linear combinations of the row vectors of A.

$.1 a_1^T + .3 a_2^T + \dots$

$$S = \{v \in R^n : v = \sum_{i=1}^{n} \alpha_i a_i \text{ where } 0 \le \alpha_i \le 1, i = 1 \dots n\}$$

Example

$$\begin{array}{cc} & \begin{array}{cc} a_1 & a_2 \end{array} \\ A = & \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \end{array}$$



The absolute value of the determinant is the volume of S

Some intuition

1) Determinant of an identity is 1, $|I| = 1$. Geometrically the volume of a unit hypercube is 1.

2) Given a matrix $A \in R^{n \times n}$, if we multiply a single row in A by a scalar $t \in R$, then the $|\hat{A}| = t|A|$.

3) Exchanging any two rows a_i^T and a_j^T of A, then the determinant of the new matrix is $|-A|$.

More properties

- a) $A \in R^{n \times n}$, $|A| = |A^T|$

b) $A, B \in R^{n \times n}$, $|AB| = |A||B|$

c) $A \in R^{n \times n}$, $|A| = 0$ iff $\underbrace{A \text{ is singular}}_{\text{question}}$ ("flat sheet")

not full rank

d) $A \in R^{n \times n}$, A is non-singular $|A^{-1}| = 1/|A|$

Determinant for 2x2 : $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

Quadratic Forms and Positive Semidefinite Matrices

Let $A \in R^{n \times n}$, vector $x \in R^n$

the scalar value

Side note

$\text{cov}[X,Y] \triangleq E[(X-E[X])(Y-E[Y])]$

$= E[XY] - E[X]E[Y]$

symmetric positive definite

$$\underbrace{x^T A x}_{\text{multivariate Gaussian}} \Rightarrow \text{quadratic form}$$

convex optimization

This results in the definition of positive semidefinite matrices.

symmetric matrix A is positive definite if for all non zero vectors $x \in R^n$, $x^T A x > 0$.

" " positive semi-definite

if " " $x \in R^n$, $x^T A x \geq 0$.

1) All positive definite matrices are full rank.

2) Given any matrix $A \in R^{m \times n}$, the matrix $G = A^T A$ (Gram matrix)

is always positive semidefinite. If $m \geq n$, then G is positive definite.

3) A matrix is positive definite if all its eigenvalues are > 0.

Eigenvalues and Eigenvectors

think ; recap...
what does a matrix do?
it acts on a vector x
performs some kind of mapping and returns to us some other point in that space.

$Ax \rightsquigarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Rightarrow$ A maps it into some other space $\Rightarrow b$

Ax parallel to x
[Eigenvectors]

$\boxed{Ax = \lambda x}$ $\quad x \neq 0$

This means that multiplying A by the vector x results in a new vector that points in the same direction as x, but scaled by λ.

For any eigenvector $x \in \mathbb{C}^n$ and scalar $t \in \mathbb{C}$, $A(cx) = cAx = \lambda \underbrace{cx}$

also an eigen vector

Rewrite the equation above:

$$(\lambda I - A)x = 0 \quad x \neq 0.$$

only if $\underbrace{|\lambda I - A|}_{\text{determinant}} = 0$

Hessians and Jacobians?
second first

second order partial derivatives

$f''(x) < 0$
local max

$f''(x) > 0$
local min

Properties

1) $Tr(A) = \sum_{i=1}^{n} \lambda_i$

2) $|A| = \prod_{i=1}^{n} \lambda_i$

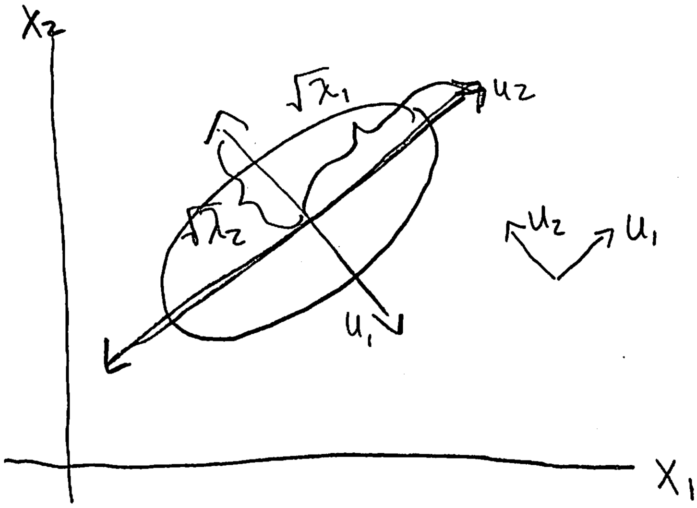
3) $Rank(A) = \sum_{i=1}^{n} \delta(\lambda_i > 0)$

4) If A is invertible / non-singular then $1/\lambda_i$ is an eigenvalue of A^{-1} with associated eigenvector x_i.

5) Eigenvalues of a diagonal matrix are just the diagonal entries.

Relationship to multivariate Gaussians

$$\mathcal{N}(x|\mu,\Sigma) \triangleq \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$



Two-Dimensional Gaussian Density

The first two eigenvectors of our covariance matrix is $\{u_1, u_2\}$

$$\Sigma = U \Lambda U^T$$

← diagonal matrix of eigenvalues

↑ orthonormal matrix of eigenvectors

For symmetric matrices the eigenvectors are orthonormal.

Mercer Kernels

Kernel function is a real-valued function of 2 arguments $K(x, x') \in \mathbb{R}$ for $x, x' \in X$.

Gaussian Kernel $K(x,x') = \exp\left(-\frac{1}{2}(x-x')^T\Sigma^{-1}(x-x')\right)$

Simplest $K(x,x') = \underbrace{x^T x'}_{\text{inner product}}$

If Gram matrix is positive definite, we can compute an eigenvector decomposition of it as follows:

$$K = U^T \Lambda U$$

Λ: diagonal eigenvalues

$$K_{ij} = \underbrace{(\Lambda^{1/2} U_{:i})^T}_{\phi(x_i)^T} \underbrace{(\Lambda^{1/2} U_{:j})}_{\phi(x_j)}$$

We call such a positive definite matrix a Mercer Kernel

Mercer's Theorem: Let $K(x,x')$ be given. Then K is a Mercer Kernel (i.e. $\exists \phi$ s.t. $K(x,x') = \phi(x)^T\phi(x)$) iff for all $\{x^{(1)}, \ldots, x^{(m)}\}$ $(m < \infty)$ the kernel matrix K is symmetric positive semidefinite. Converse holds true.

Worked out example for the eigen vector

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = 0$$

$$\Downarrow$$

$$\underbrace{\lambda^2 - 4\lambda + 3 = 0}$$

Solutions for this root are $\lambda = 1$, $\lambda = 3$

$$\text{for } \lambda = 3 \; \underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$\left.\begin{aligned} 2x_1 + x_2 &= 3x_1 \\ x_1 + 2x_2 &= 3x_2 \end{aligned}\right\} \text{ the linear equation that solves this is } x_1 = x_2$$

$$\text{for } \lambda = 1 \left\{ \begin{aligned} 2x_1 + x_2 &= x_1 \\ x_1 + 2x_2 &= x_2 \end{aligned} \right\} \text{ the linear equation for this is } x_2 = -x_1$$

Any eigenvector for $\lambda = 3$ of the form $x_1 = x_2$ will be sufficient.

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

<u>Determinant Check</u>

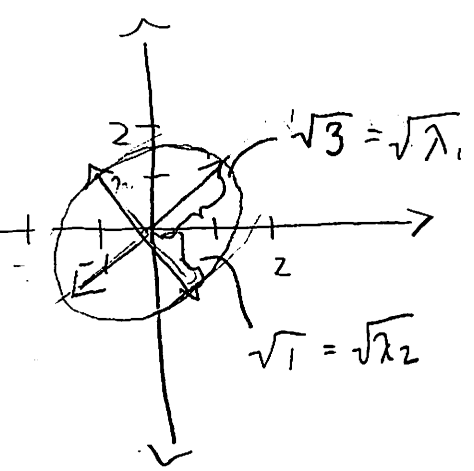
$\lambda_1 \times \lambda_3 = 3$

$|A| = 4 - 1 = 3$ ✓

<u>Trace Check</u>

$tr(A) = 2 + 2 = 4$

$\lambda_1 + \lambda_3 = 4$ ✓



<u>Determinant Trick for Σ</u>

Grab eigenvalues of Σ

$$\sum_{i=1}^{n} \log \lambda_i = \log|\Sigma|$$