Background

By: Darren Homrighausen

Background

- We will write vectors as

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

  We write this as $z \in \mathbb{R}^n$, which is "z is a member of ar-en."
- We commonly will need to "turn" the vector, which we write as

$$z^\top = \begin{bmatrix} z_1 & z_2 & \dots & z_n \end{bmatrix}$$

Necessary Background: Addition and Multiplication

We will need to extend the ideas of addition and multiplication of numbers to higher dimensional objects (vectors and matrices)

- Suppose $u, v \in \mathbb{R}^q$. Then we write u "times" v as

$$u^\top v = \sum_{i=1}^{q} u_i v_i$$

- This is known as the inner product of u and v

Necessary Background: Addition and Multiplication

- Also, for matrices $\mathbb{A} \in \mathbb{R}^{n \times p}, \mathbb{B} \in \mathbb{R}^{p \times r}$,

  (We will refer to the entry in the i^{th} row, j^{th} column of a matrix $\mathbb{A}$ as A_{ij})

$$\mathbb{A} \cdot \mathbb{B} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ & & \vdots & \\ A_{n1} & A_{n2} & \dots & A_{np} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1r} \\ B_{21} & B_{22} & \dots & B_{2r} \\ & & \vdots & \\ B_{p1} & B_{n2} & \dots & B_{pr} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{p} A_{1j} B_{j1} & \sum_{j=1}^{p} A_{1j} B_{j2} & \dots & \sum_{j=1}^{p} A_{1j} B_{jr} \\ \sum_{j=1}^{p} A_{2j} B_{j1} & \sum_{j=1}^{p} A_{2j} B_{j2} & \dots & \sum_{j=1}^{p} A_{2j} B_{jr} \\ & & \vdots & \\ \sum_{j=1}^{p} A_{nj} B_{j1} & \sum_{j=1}^{p} A_{nj} B_{j2} & \dots & \sum_{j=1}^{p} A_{nj} B_{jr} \end{bmatrix} \in \mathbb{R}^{n \times r}$$

(Often, we will omit the $\cdot$ for matrix multiplication)

Probability

What's a random variable?

Let X be a random variable. That is, X...

- Has a probability density function p_X such that the probability (denote this by $\mathbb{P}$) that X takes on a set of values A is given by[1]

$$\mathbb{P}(A) = \int_A p_X(x)dx$$

- And p_X has certain properties such as $p_X \geq 0$ and $\int p_X = 1$.

[1] Anyone who has studied probability would have serious problems with this statement. If this is you, don't quibble; we're trying to avoid unnecessary complications.

What are the properties of a random variable?

In this class, we really only care about X's

- mean (alternatively known as its expectation)

  (This is all about finding its center)

- and variance.

  (This is all about finding its spread)

What's expectation?

Imagine taking a metal rod of a certain mass.

However, its mass isn't necessarily even along its length.

Attempt to balance the rod on your finger. The balancing point is the center of mass of the rod.

Figure: A family calculates expectations

What's expectation?

Crucial connection: If we think about the density of the random variable determining where the rod's mass is distributed, then the "center of mass" is the expectation.

$$\mathbb{E}[X] = \int x \, p_X(x) dx$$

What's variance?

For variance, I'll just give you the definition

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

In words:

"variance is the average squared deviation from the average"

Note: $$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Can you take me higher?

Implicitly, we were assuming that $X \in \mathbb{R}$.

What happens if $X \in \mathbb{R}^p$?

The expectation is going to look the same, but be a vector

$$\mathbb{E}[X] \in \mathbb{R}^p$$

For variance, we need to use some matrix notation:

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] \in \mathbb{R}^{p \times p}$$

If you write this out, you'll see that this a matrix with

- The variances of the components on the diagonal
- The covariances of any two components in the off-diagonal entries.

Combine matrices and probability

We will commonly combine matrix multiplication with probability statements

Suppose that $Y \in \mathbb{R}^n$ is a random variable such that $\mathbb{E}[Y] = \mu$ and $\mathbb{V}[Y] = \Sigma$.

What is the distribution of $\mathbb{A}Y$?

It turns out expectation is linear and hence we can rearrange '$\mathbb{E}$' and '$\mathbb{A}$'

$$\mathbb{E}[\mathbb{A}Y] = \mathbb{A}\mathbb{E}[Y] = \mathbb{A}\mu$$

Variance is a little more complicated, but not much

$$\mathbb{V}[\mathbb{A}Y] = \mathbb{A}\mathbb{V}[Y]\mathbb{A}^\top = \mathbb{A}\Sigma\mathbb{A}^\top$$

Necessary Background: Notation

We will concatenate the explanatory variables into the design matrix $\mathbb{X}$, where

$$\mathbb{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} = \begin{bmatrix} X_1^\top \\ X_2^\top \\ \vdots \\ X_n^\top \end{bmatrix} \in \mathbb{R}^{n \times p}$$

In words: The features (columns) will be lower case letters and the observations (rows) will be upper case letters

We will refer to the entry in the i^{th} row, j^{th} column as X_{ij}

Necessary Background: Lengths

We will need to measure the size of both vectors and matrices.

The most common is the one we use every day Euclidean distance

(Think: the Pythagorean theorem)

$$||x||_2 = \sqrt{\sum_{k=1}^{p} x_k^2} = \sqrt{x^\top x}$$

We call this a norm and refer to this as the "ell two norm"

(Often, it will be written as squared for convenience: $||x||_2^2 = x^\top x$)

Additionally, we will need the Manhattan distance

$$||x||_1 = \sum_{k=1}^{p} |x_k|$$

We call this the "ell one norm"

Singular Value Decomposition (SVD)

SVD

A huge amount of statistics depends on (numerical) linear algebra concepts

Many, many topics in (numerical) linear algebra are implicitly motivated by the singular value decomposition (SVD)

The SVD is a generalization of the eigenvector decomposition

Instead of

$$\mathbb{X} = UDU^\top \longleftarrow \text{eigenvector decomposition}$$

we get

$$\mathbb{X} = UDV^\top \longleftarrow \text{singular value decomposition}$$

This change makes the (unique) SVD always exist

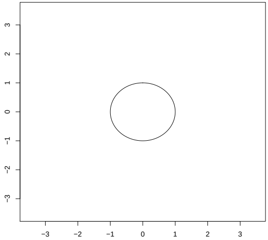
SVD

It turns out we can think of matrix multiplication in terms of circles and ellipsoids

Take a matrix $\mathbb{X}$ and let's look at the set of vectors

$$B = \{\beta : ||\beta||_2 \leq 1\}$$

This is a circle!

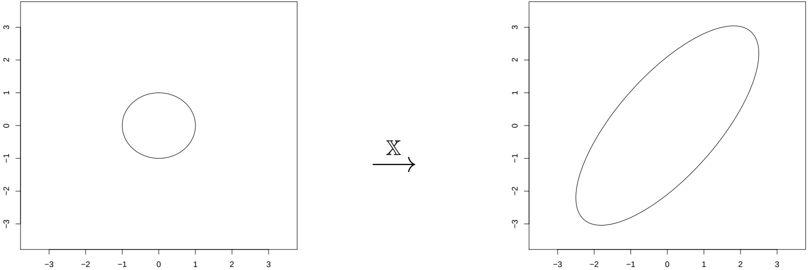


SVD

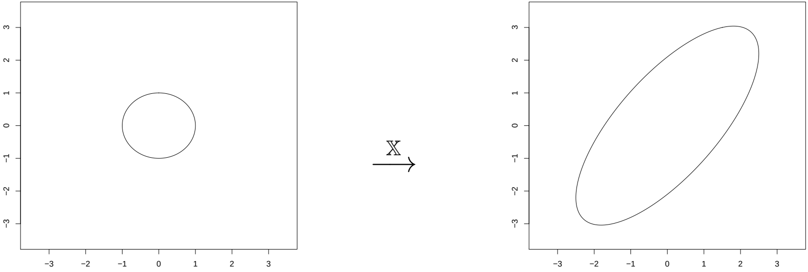
What happens when we multiply vectors in this circle by $\mathbb{X}$?

Let

$$\mathbb{X} = \begin{bmatrix} 2.0 & 0.5 \\ 1.5 & 3.0 \end{bmatrix} \text{ and } \mathbb{X}\beta = \begin{bmatrix} 2\beta_1 + 0.5\beta_2 \\ 1.5\beta_1 + 3\beta_2 \end{bmatrix}$$



SVD

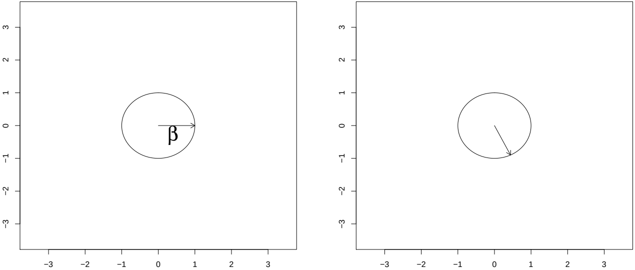


What happened?

1. The coodinate axis gets rotated
2. The new axis gets elongated (making an ellipse)
3. This ellipse gets rotated

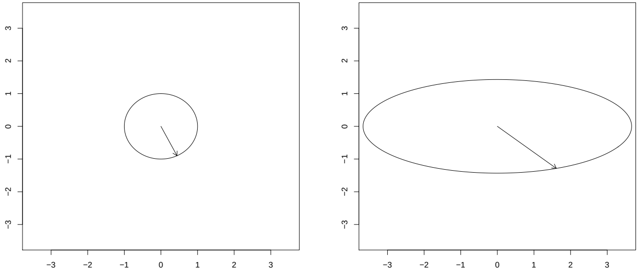
Let's break this down into parts...

SVD



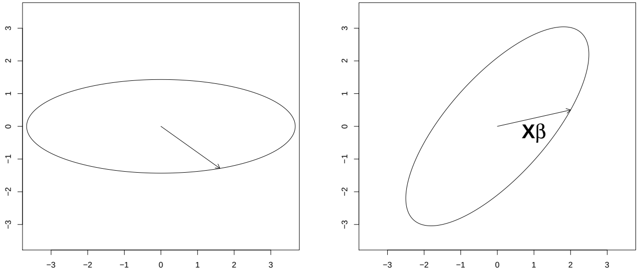
1. The coordinate axis gets rotated

SVD



1. The coordinate axis gets rotated
2. The new axis gets elongated (making an ellipse)

SVD



1. The coordinate axis gets rotated
2. The new axis gets elongated (making an ellipse)
3. This ellipse gets rotated

Necessary Background: Rotation

Rotations: These can be thought of as just reparameterizing the coordinate axis. This means that they don't change the geometry.

As the original axis was orthogonal (that is; perpendicular), the new axis must be as well.

Necessary Background: Rotation

Let v_1, v_2 be two normalized, orthogonal vectors. This means that:

$$v_1^\top v_2 = 0 \quad \text{and} \quad v_1^\top v_1 = v_2^\top v_2 = 1$$

In matrix notation, if we create V as a matrix with normalized, orthogonal vectors as columns, then:

$$V^\top V = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = I$$

Here, I is the identity matrix.

Necessary Background: Elongation

Elongation: These can be thought of as stretching vectors along the current coordinate axis. This means that they **do** change the geometry by distorting distances.

Elongations are the result of multiplication by a diagonal matrix (note: we just saw a very special case of such a matrix: the identity matrix I)

All diagonal matrices have the form:

$$D = \begin{bmatrix} d_1 & 0 & 0 & \ldots & 0 \\ 0 & d_2 & 0 & \ldots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \ldots & d_p \end{bmatrix}$$

SVD

Using this intuition, for any matrix $\mathbb{X}$ it is possible to write its SVD:

$$\mathbb{X} = UDV^\top$$

where

- U and V are orthogonal (think: rotations)
- D is diagonal (think: elongation)
- The diagonal elements of D are ordered as

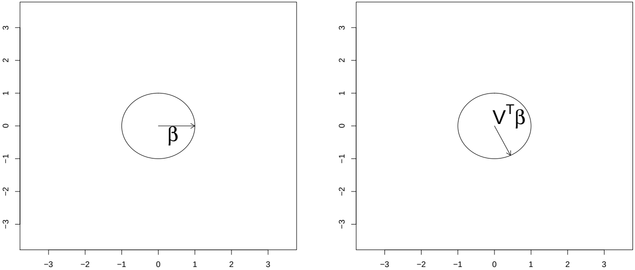
$$d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$$

Many properties of matrices can be 'read off' from the SVD.

Rank: The rank of a matrix answers the question: how many dimensions does the ellipse live in? In other words, it is the number of columns of the matrix $\mathbb{X}$, not counting the columns that are 'redundant'

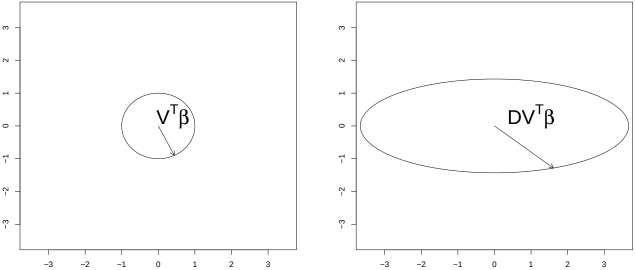
It turns out the rank is exactly the quantity q such that $d_q > 0$ and $d_{q+1} = 0$

SVD



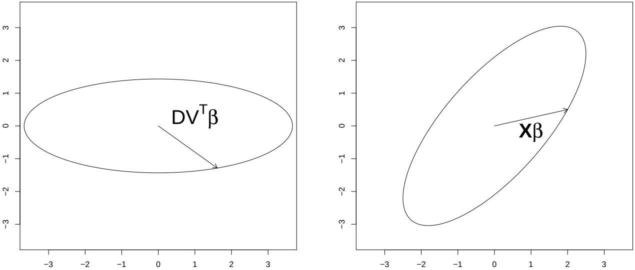
1. The coordinate axis gets rotated (Multiplication by $V^\top$)

SVD



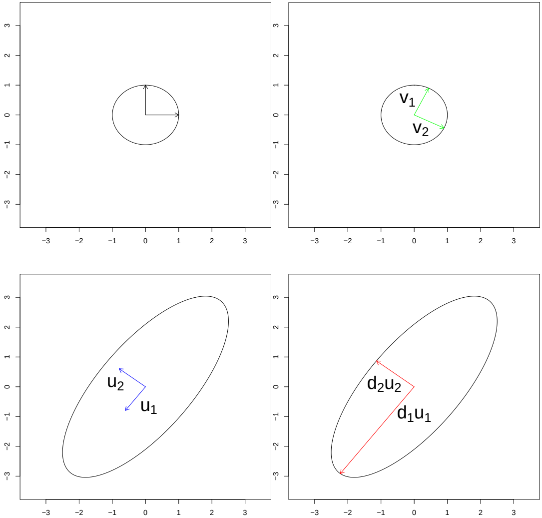
1. The coordinate axis gets rotated (Multiplication by $V^\top$)
1. The new axis gets elongated (Multiplication by D)

SVD



1. The coordinate axis gets rotated (Multiplication by $V^\top$)
1. The new axis gets elongated (Multiplication by D)
2. This ellipse gets rotated (Multiplication by U)

SVD [ONE LAST TIME]



Summary:

Of all the possible axes of the original circle, the one given by v_1, v_2 has the unique property:

$$\mathbb{X} v_j = d_j u_j$$

for all j.

Lastly:

$$\mathbb{X} = \sum_j d_j u_j v_j^\top$$