Introduction, Notation, and Overview

-Applied Multivariate Analysis-

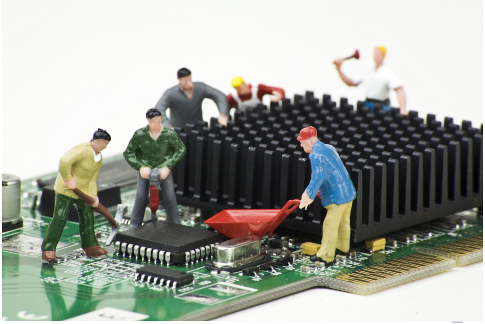
Lecturer: Darren Homrighausen, PhD

Class Overview

Over the next semester we will cover (my opinion of) Applied Multivariate analysis.

This can be described by the more modern titles

- Data mining
- Statistical learning
- Data analytics



Class Overview

Practically speaking, this means:

- Finding relationships between a group of explanatory and response variables that provides good predictive performance
- Reducing the size of the group of variables for scientific, statistical, or computational purposes

and, perhaps most importantly..

- Knowing the techniques, how they work, when they apply, and how to implement them

Necessary Background: Notation

- We will write vectors as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

  We write this as $x \in \mathbb{R}^n$, which is "x is a member of ar-en."

- We commonly will need to "turn" the vector, which we write as

$$x^\top = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

  Here, the superscript "T" takes a vector and flips it on its side.

Necessary Background: Notation

If we have two vectors, we will double subscript them

Suppose $x_1, x_2 \in \mathbb{R}^n$, then

$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \quad \text{and} \quad x_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix}$$

Necessary Background: Notation

Often, we will combine many vectors into a matrix

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

As much as possible

- Lower case Roman letters will be columns
- Upper case Roman letters will be rows

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 $x_1, x_2 \in \mathbb{R}^q$. Then we write "$x_1$ times x_2" as

$$x_1 \cdot x_2 = \sum_{j=1}^{q} x_{1j} x_{2j} = x_1^\top x_2$$

Necessary Background: Addition and Multiplication

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

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

$$= \begin{bmatrix} A_1^\top b_1 & A_1^\top b_2 & \dots & A_1^\top b_r \\ A_2^\top b_1 & A_2^\top b_2 & \dots & A_2^\top b_r \\ & & \vdots & \\ A_n^\top b_1 & A_n^\top b_2 & \dots & A_n^\top b_r \end{bmatrix} \in \mathbb{R}^{n\times r}$$

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

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)

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

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

Additionally, we will need the Manhattan distance

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

We call this the "ell one norm"

Necessary Background: Lengths

For matrices, we will just define something very related to 'length'
(but it doesn't technically qualify)

Many times, we are interested in the size of the diagonal of a matrix

This is known as the trace and is defined to be

$$\text{trace}(\mathbb{X}) = \sum_{j=1}^{p} \mathbb{X}_{jj}$$

That is, the trace is the sum of the diagonal entries.

Singular Value Decomposition (SVD)

Necessary Background: SVD

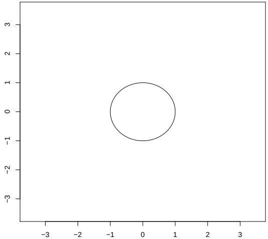
It turns out we can think of matrix multiplication in terms of circles and ellipses

(The plural is technically ellipsoids, but this term seems to freak people out)

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

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

This is a circle!

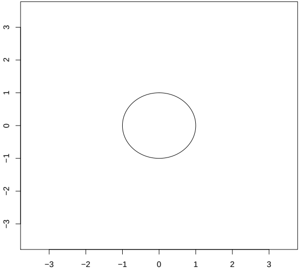


Necessary Background: SVD

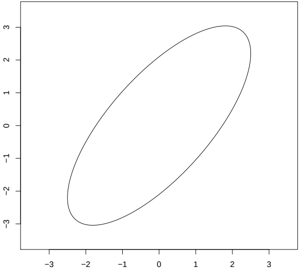
What happens when we multiply vectors in this circle by $\mathbb{X}$?

Let

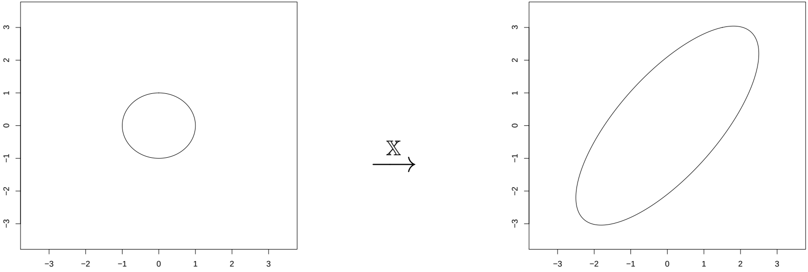
$$\mathbb{X} = \begin{bmatrix} 2.0 & 0.5 \\ 1.5 & 3.0 \end{bmatrix} \text{ and } \mathbb{X}\beta = \begin{bmatrix} 2\beta_1 + 0.5\beta_2 \\ 1.5\beta_1 + 3\beta_2 \end{bmatrix}$$



$\xrightarrow{\mathbb{X}}$



Necessary Background: SVD

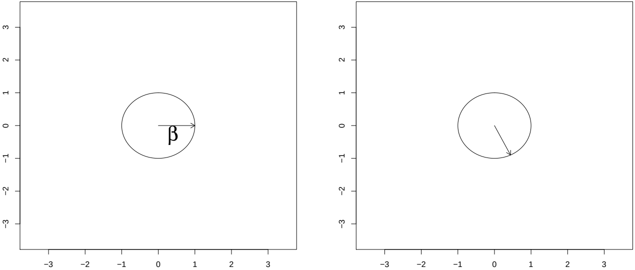


What happened?

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

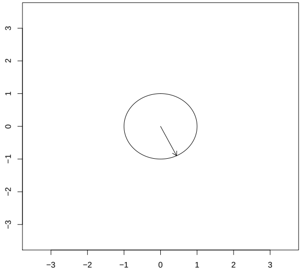
Let's break this down into parts...

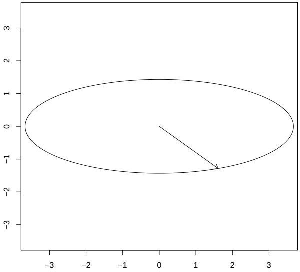
Necessary Background: SVD



1. The coordinate axis gets rotated

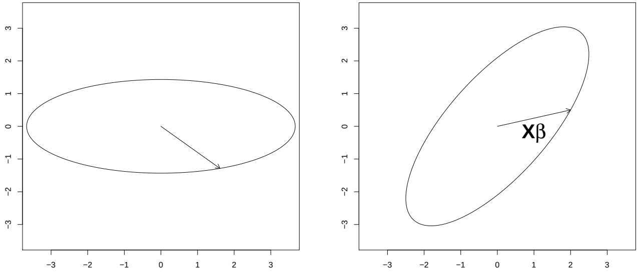
Necessary Background: SVD





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

Necessary Background: SVD



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

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 $\mathbf{v}_1, \mathbf{v}_2$ be two normalized, orthogonal vectors. This means that:

$$\mathbf{v}_1^\top \mathbf{v}_2 = 0 \quad \text{and} \quad \mathbf{v}_1^\top \mathbf{v}_1 = \mathbf{v}_2^\top \mathbf{v}_2 = 1$$

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:

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

Necessary Background: 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

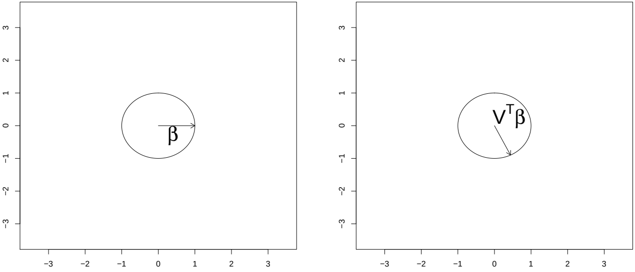
$$d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$$

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'

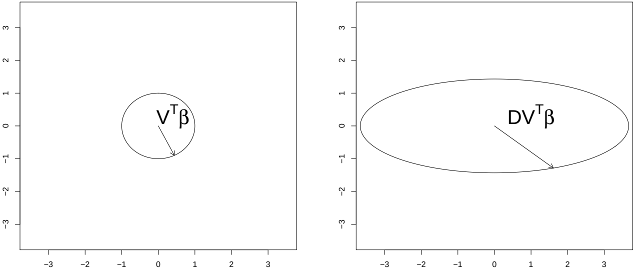
It turns out the rank is exactly the quantity q such that $d_q > 0$ and $d_{q+1} = 0$

Necessary Background: SVD



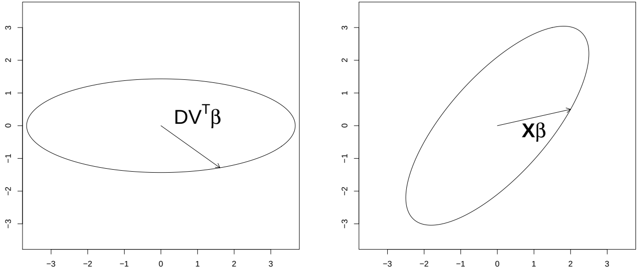
1. The coordinate axis gets rotated (Multiplication by $V^\top$)

Necessary Background: SVD



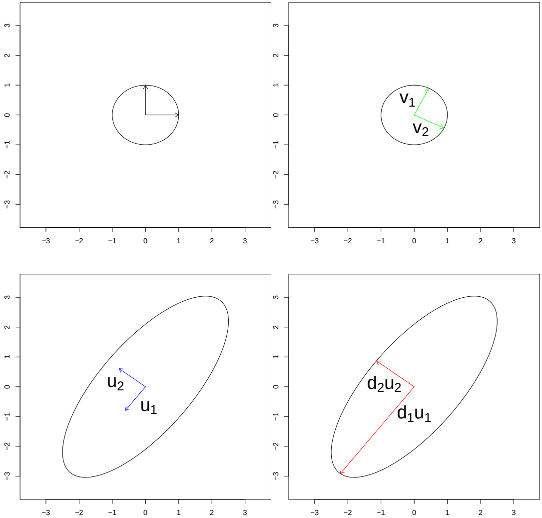
1. The coordinate axis gets rotated (Multiplication by $V^\top$)
1. The new axis gets elongated (Multiplication by D)

Necessary Background: 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)

Necessary Background: 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:

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

Probability

Necessary Background: Probability

For this class, we really don't need to know too much probability.

Again, I would be satisfied with you accepting that certain manipulations are reasonable, rather than 'understanding' everything.

That being said, let's take it away...

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 $\mathbb{X}Y$?

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

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

Variance is little more complicated, but not much

$$\mathbb{V}[\mathbb{X}Y] = \mathbb{X}\mathbb{V}[Y]\mathbb{X}^\top = \mathbb{X}\Sigma\mathbb{X}^\top \qquad \text{(check this!)}$$

I thought this was a statistics class...