Introduction to Mathematics for Political Science: Inner Product Spaces, Orthogonality, Projection

4 September 2018

Suppose we observed m features related to development across a sample of n countries – GDP per capita, life expectancy, literacy rate, etc. We're interested in creating some "index of development" that describes these data with a single number. These observations can be thought of as elements of a set. We started our study of linear spaces with a motivating question. How can we summarize many statistics related to a country's development with a single number, an "index of development"? We learned in the last lecture that a linear space with an inner product is an inner product space. Today, we put inner product spaces to work in order to construct our index of development.

Orthogonality

You studied orthogonality when learning about vectors and matrices. Two vectors are *orthogonal* if their inner product is zero. Geometrically, vectors are orthogonal when they are perpendicular, making a 90 degree angle with one another. The process of dimension reduction necessary to generate our Development Index is really about finding a set of orthogonal linear subspaces that best "fit" the data, and defining our data in terms of those subspaces. This will make more sense once we understand better orthogonality and projection.

Definition: Two *vectors* x and y are *orthogonal* iff $x^T y = 0$.

Theorem (Pythagorus): $x^T y = 0 \implies \|x\|^2 + \|y\|^2 = \|x+y\|^2$

Definition: Two *spaces* X and Y are orthogonal if for all $x \in X$ and $y \in Y$, $x^T y = 0$.

Definition: An *orthogonal basis*[1] for a linear space X is a set of basis vectors $S = \{x_1, ..., x_n\}$ such that for all basis vectors x_i and x_j with $i \neq j$, $x_i^T x_j = 0$.
   Remark: The standard basis is an orthogonal basis.
   Example: Name a few orthogonal bases for $\mathbb{R}^2$.
   Definition: An *orthonormal basis* is an orthogonal basis composed entirely of unit vectors.

[1] Remember, a *basis* for a linear space X is a set $y_1, ..., y_n \in S$ such that for all $x \in X$,

$$\alpha_1 y_1 + ... + \alpha_n y_n = x$$

Projection

Now we get to put all these concepts to use, and start thinking about how to move our data between different linear subspaces. The idea

behind *projection* is to take an element of a linear space, and to push it into some linear subspace of that space.

Let S be a subspace of a normed linear space X and take some $x \in X$. Let $p_s : X \to S$ be a function that maps this $x \in \mathbb{R}^n$ to its nearest point in S. Formally,[2]

$$d(x, p_s(x)) = d(x, S) = \min \{d(x, y) | y \in S\}$$

How are orthogonality and projection related? Assume we're in the familiar metric space of $(\mathbb{R}^n, d_2)$. If $x - p_s(x)$ is orthogonal to S, then $p_s(x)$ must be a projection (and vice versa). To see why, take some other $y \neq p_x(x) \in S$. We have

$$\|x - p_s(x)\|^2 \leq \|x - p_s(x)\|^2 + \|p_s(x) - y\|^2$$

Since $p_s(x) - y \in S$ and $x - p_s(x) \perp S$, Pythagorus's Theorem applies

$$\begin{aligned} \|x - p_s(x)\|^2 &\leq \|x - p_s(x)\|^2 + \|p_s(x) - y\|^2 \\ &= \|x - p_s(x) + p_s(x) - y\|^2 \qquad \text{Pythagorus} \\ &= \|x - y\|^2 \end{aligned}$$

Then,

$$\|x - p_s(x)\| \leq \|x - y\|$$

allowing us to conclude $d_2(x, p_s(x)) \leq d_2(x, y)$ for all $y \in S$.

This formal definition helps us see why orthogonality plays such a crucial role in projection. If $x - Px$ is orthogonal to S, we know that we're taking the shortest path from x to S, satisfying the requirements of a projection.[3]

Example: Consider the vector $a = (1, 2, 3)$. This vector occupies $\mathbb{R}^3$. Suppose we wanted to *project* this vector onto the z axis, a one dimensional subspace of $\mathbb{R}^3$. To do so, we simply multiply by the matrix

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

which takes third component of a and drop the other two components, leaving $Pa = p_3 = (0, 0, 3)$. Suppose we wanted to project into the xy-plane? Here, we simply let,

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

dropping the z component, and giving $Pa = p_{12} = (1, 2)$ What's special about P and p?

1. Projection matrices are *symmetric*: $P^T = P$

2. Projection matrices are *idempotent*: $PP = P$

3. Projections p are stable: $Pp = p$[4]

For empirical political scientists this is not an academic conversation. Return to the cross national data discussed at the beginning of the lecture. We observe N countries and M features of those countries. Suppose we wanted to know how life expectancy varies with literacy rates and per capita gdp. We can store life expectancy observations in a vector $y_{1 \times N}$. We can store the other observations in a matrix $A_{N \times 2}$. In a simple linear model, we're interested in finding $\beta_{1 \times 2}$ that relates per capita gdp and literacy to life expectancy. So we could attempt to solve

$$y = \beta A^T$$

With $N > M - 1$, this system is unlikely to have a solution unless the data exhibit severe collinearity. We solve this problem through projection. We'll explore this in more detail in tomorrow's lecture. But you're now equipped with the tools to think about *how* to solve it. β defines a linear subspace of the N-dimensional space of rows. If we *project y* into this subspace, then we can take linear combinations of the columns of A and reach y.[5]

Let's practice with a simple example.

Example: Now, suppose we wanted and orthogonal projection of an element in two dimensional space onto some line passing through the origin.[6] Call this line the vector a and the element with the vector b. We know the projection will lie along this line

$$p = \hat{x}a$$

where $\hat{x}$ is some scalar. Let e represent the vector connecting p and a

$$e = b - \hat{x}a$$

We know from our definition of an orthogonal projection that this vector is orthogonal to a, or

$$(a)^T(b - \hat{x}a) = 0$$

Now apply your linear algebra skills to solve for $\hat{x}$. We see

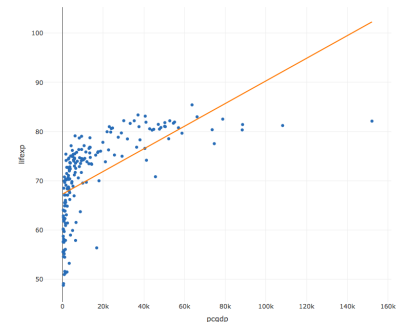$$\hat{x} = \frac{a^T b}{a^T a}$$

Our projection is

$$p = \hat{x}a = \frac{a^T b}{a^T a}a$$

What does this tell us about the projection matrix P?

Figure[7] demonstrates this algebra and each component of the projection visually.

[4] Verify these properties for the example above.

[5] Think about this discussion in the context of a simple 2-dimensional linear regression.

To where are the observations in y projected?

[6] For the more general case of projection onto a subspace, see Strang p. 209. This exercise should give you the intuition for this case.
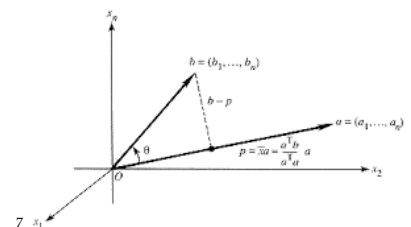
Application: Principal Component Analysis (PCA)

Principal component analysis (PCA) is a method used to "discover" underlying dimensions in data. Using the language we've learned today, this method iteratively finds orthogonal one-dimensional *linear subspaces* that best explain the variance in the data, or minimize the distance between these subspaces and the elements of the data. We can think of this as choosing a "most explanatory" *orthogonal basis* for our data. The *projections* of these data into these linear subspaces are called component scores and are interpreted in relation to the underlying data.

We can think of the problem of developing an Development Index as a constrained prediction problem. If I asked you to predict the economic, health, and education outcomes of a country on the basis of a single number, one for each country, what number would do the best job? Arguably, this number would be a good candidate for our development index, and principal components analysis provides us with a tool to recover that number.

Implementing PCA requires an understanding of eigenvalues and matrix decompositions – material you'll be exposed to in later courses. Still, we can understand what PCA does using concepts we've studied today. Orthogonal matrices are sometimes called *rotation matrices*. This is because as linear transformations, they preserve the geometry of the inputs they act upon. In 2 dimensions, rotation matrices are given by[8]

[8] confirm for yourself that this matrix is orthogonal.

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

Suppose we were analyzing a data set consisting of three observations, $x_1 = (-1, -1)$, $x_2 = (0, 0)$, $x_3 = (1, 1)$. While these data points lie in $\mathbb{R}^2$, it is clear that they reside in a linear subspace of $\mathbb{R}^2$, namely the 45 degree line. Rotating these data by -45 degrees would place each observation on the x-axis. To see this, store the data in the matrix

$$X_{M \times N} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

and consider the rotation matrix that implements a -45 degree rotation

$$R_{-45} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Applying this rotation to our data matrix X gives

$$Y = R_{-45}X = \begin{pmatrix} \sqrt{2} & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

where it is immediately clear that Y is rank 1 and the rows of R_{-45} represent our new matrix.

More generally, PCA selects an n-dimensional *orthogonal basis*, rotating and scaling our standard coordinate system to better fit the data, and then projects these data into the orthogonal basis. The rows of the orthogonal basis are referred to as "component scores." You can see this in action in Figure[9].

I grabbed some data on our outcomes of interest from the World Bank. Life expectancy, GDP per capita, and literacy rates, and conducted principal components analysis on them. The component scores for the first dimension are visualized in Figure 4.

Economic Development Index
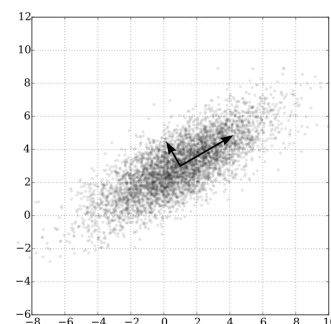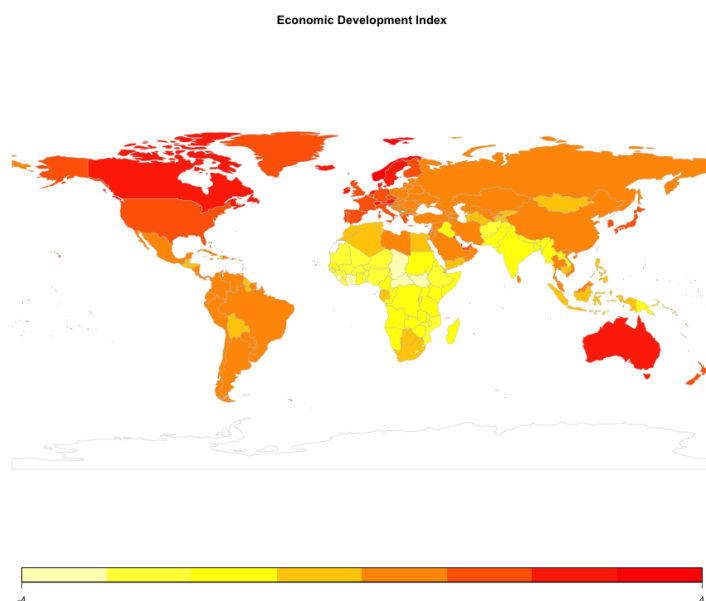
[9]

References

- Moore, Will H. and David A. Siegel. *A Mathematics Course for Political and Social Research*. Chapter 13.
- Strang, Gilbert. *Introduction to Linear Algebra*. Chapters 4, 6.
- Carter, Michael. *Foundations of Mathematical Economics*. Chapter 3.4, 3.6.
- Ok, Efe A. *Real Analysis with Economic Applications*. Chapter D.
- Shlens, Jon. "A Tutorial on Principal Component Analysis"