Linear Algebra: Relation to Statistics

Daven Farnham

February - May 2016

1 Correlation

A very basic way to think of correlation is to consider a data set of, let's say, weight and life expectancy:

weight	age
200lbs	78
170lbs	75
250lbs	62
185lbs	81
225lbs	68

the average life expectancy from this (admittedly small) sample is 72.8 while the average weight is 206. Let's construct 2 vectors that represent the difference of each individual value from the mean:

$$w = \text{greater than average} = \begin{bmatrix} -6 \\ -36 \\ 44 \\ -21 \\ 19 \end{bmatrix}, a = \text{less than average} = \begin{bmatrix} -5.2 \\ -2.2 \\ 10.8 \\ -8.2 \\ 4.8 \end{bmatrix}$$

If you then take the dot product of these two vectors $w \cdot a$, you'll get a positive value (849), which indicates there's a positive correlation between having a weight greater than average and living less than the average, i.e. obesity and a shorter life expectancy. In other words, if your weight is less than the mean 206, you're more likely to live longer than those people weighing (> 206).

2 Best Fit

A good way to do this is to use least-squares approximation. Let's say you have a matrix, A:

$$\begin{bmatrix} 2 & 0 \\ 0 & 4 \\ 1 & 3 \end{bmatrix}$$

this is clearly a plane in two dimensions, with kernel = $\begin{bmatrix} 2 & 3 & -4 \end{bmatrix}$. But, let's say I give you a vector that can't be constructed from a linear combination of A's column vectors, say $\vec{b} = \begin{bmatrix} 3 & 5 & 8 \end{bmatrix}$. No vector $\vec{x}$ in $A\vec{x}$ will equal $\vec{b}$. Instead, find the best fit, i.e. find the coefficients to the linear combination of column vectors ($\vec{x}^*$) that'll give you $\text{proj}_A(\vec{b})$. There are two ways to do this. You can either:

- Find the projection of $\vec{b}$ onto A, let's call it $\vec{p}$.
- Solve the augmented matrix where x^* = A|p.

Or, you can use the fact that the space orthogonal to the Image(A), $A^\perp$, is in the kernel of A^T. Knowing this:

$$A^T(\vec{b} - Ax^*) = 0$$
$$A^T\vec{b} - A^TAx^* = 0$$
$$A^T\vec{b} = A^TAx^*$$
$$(A^TA)^{-1}A^T\vec{b} = x^*$$

This method is nice since you know that A^TA will be invertible; it creates a symmetric matrix and all symmetric matrices are not only invertible, but also have an orthogonal eigenbasis according to the Spectral Theorem.

See python notebook Approximations.ipynb.

3 Eigenvectors

An eigenvector is defined as a vector $\vec{x}$ s.t. for a given matrix A, $A\vec{x} = \lambda\vec{x}$ where λ is a constant. So, for example, given a matrix representing the orthogonal projection onto a plane $4x + 5y - 6z = 0$, any vector within the plane will be a valid eigenvector with eigenvalue 1, while the vector perpendicular to the plane will be an eigenvector with eigenvalue 0. A possible eigenbasis for the space, therefore, could be something like:

$$\begin{bmatrix} 3 & 0 & 4 \\ 0 & 6 & 5 \\ 2 & 5 & -6 \end{bmatrix}$$

What's useful about eigenvectors and eigenvalues is that they allow you to create closed equations for future values. If you want to calculate $A.A.A.A.A\vec{x}$ or, put more simply, $A^5\vec{x}$, you'd ideally like to avoid having to multiply the matrix A 5 times. If I can construct $\vec{x}$ through a linear combination of eigenvectors $\vec{v_1}, \vec{v_2}, ...\vec{v_n}$ (i.e. there's an eigenbasis for the space), I can write $\vec{x}$:

$$\vec{x} = c_1\vec{v_1} + c_2\vec{v_2} + ... + c_n\vec{v_n}$$
$$A^5\vec{x} = c_1\lambda_1^5\vec{v_1} + c_2\lambda_2^5\vec{v_2} + ... + c_n\lambda_2^5\vec{v_n}$$

3.1 Finding Eigenvectors

Let's say you want to find the eigenvectors for a given matrix A:

$$A\vec{x} = \lambda\vec{x}$$

$$A\vec{x} - \lambda\vec{x} = 0$$

$$A\vec{x} - I\lambda\vec{x} = 0$$

$$(A - I\lambda)\vec{x} = 0$$

$\vec{x}$ is thus in the kernel of $(A - I\lambda)$. This means that, assuming there are eigenvectors and the kernel doesn't equal $\vec{0}$, $\mathrm{Det}(A - I\lambda) = 0$. Solving the characteristic polynomial will give the eigenvalues; the kernel of the subsequent matrix $(A - I\lambda)$, plugging in for λ, will give you the corresponding eigenvectors.

3.2 Fibonacci Sequence and the Golden Ratio

The Fibonacci sequence = (0, 1, 1, 2, 3, 5, 8, 13, ...). It's clear that for positions $x > 2$, $value_{x+1} > value_x$. But, as x approaches infinity, what's the ratio of these values? Does $value_{x+1}$ get increasingly large compared to $value_x$? Or, does this ratio approach a constant value?

Create a matrix A that represents the values in the sequence:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_2 \\ v_1 \end{bmatrix} = \begin{bmatrix} v_2 + v_1 \\ v_2 \end{bmatrix}$$

Given the initial vector $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, A$\vec{i} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, A$^2\vec{i} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, etc.

Let's find the eigenvalues for the above matrix A:

$$(1 - \lambda)(-\lambda) - 1$$

$$\lambda^2 - \lambda - 1$$

$$\lambda_1 = \tfrac{1+\sqrt{5}}{2},\ \lambda_2 = \tfrac{1-\sqrt{5}}{2}$$

$$\vec{v_1} = \begin{bmatrix} 1 \\ \frac{-1+\sqrt{5}}{2} \end{bmatrix}$$

$$\vec{v_2} = \begin{bmatrix} 1 \\ \frac{-1-\sqrt{5}}{2} \end{bmatrix}$$

Now, let's create our initial vector through a linear combination of our eigenvectors:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1/10(5 + \sqrt{5}) \begin{bmatrix} 1 \\ \frac{-1+\sqrt{5}}{2} \end{bmatrix} + 1/10(5 - \sqrt{5}) \begin{bmatrix} 1 \\ \frac{-1-\sqrt{5}}{2} \end{bmatrix}$$

A closed equation as n approaches infinity, therefore:

$$\begin{bmatrix} value_{n+1} \\ value_n \end{bmatrix} = 1/10(5+\sqrt{5})(\tfrac{1+\sqrt{5}}{2})^n \begin{bmatrix} 1 \\ \frac{-1+\sqrt{5}}{2} \end{bmatrix} + 1/10(5-\sqrt{5})(\tfrac{-1+\sqrt{5}}{2})^n \begin{bmatrix} 1 \\ \frac{-1-\sqrt{5}}{2} \end{bmatrix}$$

$$= \frac{1/10(5+\sqrt{5})(\frac{1+\sqrt{5}}{2})^n + 1/10(5-\sqrt{5})(\frac{-1+\sqrt{5}}{2})^n}{\frac{\sqrt{5}}{5}(\frac{1+\sqrt{5}}{2})^n - \frac{\sqrt{5}}{5}(\frac{1-\sqrt{5}}{2})^n}$$

$$= \frac{1/10(5+\sqrt{5})(\frac{1+\sqrt{5}}{2})^n}{\frac{\sqrt{5}}{5}(\frac{1+\sqrt{5}}{2})^n}$$

$$= \frac{1/10(5+\sqrt{5})}{\frac{\sqrt{5}}{5}}$$

$$= \frac{1+\sqrt{5}}{2}$$

So, as you get further along the Fibonacci sequence, the ratio of the two values approaches the golden ratio.

4 Complex Eigenvectors

Any time an eigenvalue is of the form $a + bi \mid b \neq 0$, it's a complex, and not just a real, eigenvalue. Complex eigenvalues come from matrices like rotational matrices where, in the real plane, A$\vec{x}$ doesn't seem to be a simple scaling of $\vec{x}$. When graphed, complex eigenvalues will spiral **in** if $-1 < ||a + bi|| < 1$, **circle** if $||a + bi|| = 1$, and otherwise spiral **outward**.

Given the rotational matrix A:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

the eigenvalues are $\pm i$. For any vector, then, A$\vec{x}$ forms a closed circle about the origin.

5 Single Value Decomposition

Like least-squares approximation, this is another useful tool, primarily in the realm of lossy compression.

5.1 Image of Unit Circle

Let's say you're given a 2 x 2 matrix A with transformation T. You can choose two orthonormal vectors x_1 and x_2 s.t. $T(x_1)$ and $T(x_2)$ are also orthogonal. Regardless of the 2 x 2 matrix, by the Spectral Theorem, $A^T A$ will have an orthonormal eigenbasis.

$$A\vec{x_1} \cdot A\vec{x_2} = 0$$
$$\vec{x_1}^T A^T A\vec{x_2} = 0$$
$$\vec{x_1}^T \lambda_2 \vec{x_2} = 0$$
$$\vec{x_1} \cdot \lambda_2 \vec{x_2} = 0$$
$$\lambda_2(\vec{x_1} \cdot \vec{x_2}) = 0$$

Since the eigenvalue $\neq 0$ the dot product of the two vectors must, confirming they're orthogonal.

Thinking of this geometrically, the unit circle can be written as $\vec{u} = \cos(t)x_1 + \sin(t)x_2$ where t varies between 0 and 2π. The transformation, therefore, equals:

$$T(\vec{u}) = \cos(t)A(x_1) + \sin(t)A(x_2)$$

See Mathematica notebook Image of Unit Circle.nb

5.2 SVD

Given an n x m matrix A, you can write A as UΣV^T where V is the matrix made up of orthonormal eigenvectors from A^TA (V^T = V^{-1}), Σ is a diagonal of singular values, and U = $\frac{1}{\sigma_i}A\vec{v_i}$ for all vectors $v_i \in V$.

For an input vector $\vec{x}$, then, V^T changes the coordinate system to the eigenbasis, Σ rescales the vectors based on the square root of their eigenvalues, while U performs any rotation (dividing by σ to avoid scaling twice).

If you structure Σ s.t. $\sigma_1 > \sigma_2 > ... > \sigma_n$ you can truncate some of the smaller sigmas, keeping only those values that cause the greatest change.

5.2.1 Dimensions

Let's say A is a 200 x 300 matrix. A^{T}A, then, will be a 300 x 300 matrix with 90000 values. Out of this, V^T will be 300 x 300 since it contains an eigenbasis for the entire space (A^{T}A). Σ, though, will only be 200 x 300, since it can only have rank(A) nonzero values. U also depends on the rank(A); A($\vec{v_1}$) = 200 x 300 (300 x 1) = 200 X 1. Do this 200 times (for each singular value), and you get a 200 x 200 matrix.

$$\text{A} = \text{U}\Sigma\text{V}^T$$
$$\text{A} = 200 \text{ x } 300 = (200 \text{ x } 200)\ (200 \text{ x } 300)\ (300 \text{ x } 300)$$

If I want to give a rough sketch of the data, I could give you the 20 vectors associated with the most significant singular values. So now, instead of V^T being 300 x 300, it's 300 x 20, and UΣ is 200 x 20, meaning instead of 60000 values (300 x 200), I can get the gist of the transformation of A from only (300 x 20) + (200 x 20) = 10000 values.