CISC 271 Class 11

Orthogonal Projection

Text Correspondence: §4.2

Main Concepts:

- *Projection using a normal vector*
- *Scalar weight for a single vector*
- *Weight vector for a subspace*
- *Normal equation for projection*

Sample Problem, Data Analysis: For a new data vector, how do we "best" represent it as a sum of known vectors?

In this course, so far we have explored data as being in "clusters". Another important way to explore data is to seek linear relationships between data columns. For example, in PCA we learned how to produce one or more "scores" that we computed as linear sums of the data columns. One explanation of PCA is that it is a *projection* from a higher-dimensional space – spanned by the columns of the data matrix – to a lower-dimensional space that is spanned by the columns of the score matrix.

In general, one of the meanings of the word "projection" is a systematic transformation of an object to a known, or preferred, description. Usually this means taking an object of a certain dimension and describing it using a lower number of dimensions. One common usage is to transform a 3D point on a sphere – such as an approximation of the Earth – to a 2D point on a planar cartographic map.

In linear algebra, the preferred description is a basis of some ND vector subspace. The object to be transformed is a vector of higher dimensions, so we want to systematically transform the higher-D vector to a lower-D vector. In this class we will examine how to do this by finding the vector in the vector subspace that is geometrically nearest to the given vector.

11.1 Projecting to a 1D Subspace: Vector to Vector

The simplest non-trivial case is that we have a description of a 1D subspace as a single non-zero data vector, which we will write as $\vec{a}$. The problem is: given a new data vector $\vec{c}$, find a scalar multiple of $\vec{a}$ that is geometrically nearest to $\vec{a}$.

We can illustrate this problem with size 2 vectors, as shown in Figure 11.1. There is some vector $\vec{p}$ that is geometrically nearest to $\vec{c}$ such that $\vec{p}$ is a scalar multiple of $\vec{a}$; this is another way of saying that the vector $\vec{a}$ is a basis for a 1D subspace.

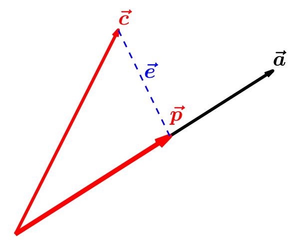

Figure 11.1: Vector $\vec{c}$ projects to $\vec{p}$, which is a scalar multiple of $\vec{a}$.

Let us formalize the problem and solve it. By "$\vec{p}$ is a scalar multiple of $\vec{a}$" we mean that there is some real number, a weight w, such that

$$\vec{p} \stackrel{\text{def}}{=} w\vec{a} \tag{11.1}$$

By "geometrically nearest" we mean that $\vec{c}$ and $\vec{p}$ are, respectively, the hypotenuse of a right triangle and one leg of the triangle. We can define an error vector $\vec{e}$ as the vector that starts at $\vec{p}$ and ends at $\vec{c}$, so

$$\vec{e} \stackrel{\text{def}}{=} \vec{c} - \vec{p} \tag{11.2}$$

The constraint of $\vec{e}$ being a leg of the right triangle, or that the $\vec{e}$ is perpendicular to $\vec{a}$, can be written as

$$\vec{a} \perp \vec{e} \quad \equiv \quad \vec{a} \cdot \vec{e} = 0 \quad \equiv \quad \vec{a}^T \vec{e} = 0 \tag{11.3}$$

Expanding the constraint of Equation 11.3 with the definitions of Equation 11.1 and Equation 11.2, we can solve for the weight w as

$$\begin{array}{rrcl}
 & \vec{a}^T\vec{e} & = & 0 \\
\Rightarrow & \vec{a}^T(\vec{c} - \vec{p}) & = & \vec{0} \\
\Rightarrow & \vec{a}^T\vec{c} - \vec{a}^T\vec{p} & = & \vec{0} \\
\Rightarrow & \vec{a}^T\vec{p} & = & \vec{a}^T\vec{c} \\
\Rightarrow & \vec{a}^T[w\vec{a}] & = & \vec{a}^T\vec{c} \\
\Rightarrow & (\vec{a}^T\vec{a})w & = & \vec{a}^T\vec{c} \\
\Rightarrow & w & = & \dfrac{\vec{a}^T\vec{c}}{\vec{a}^T\vec{a}}
\end{array} \tag{11.4}$$

Knowing the value of the scalar weight w from equation 11.4, the projection of the data vector $\vec{c}$ to the 1D subspace described by $\vec{a}$ is

$$\begin{aligned} \vec{p} &= w\vec{a} \\ &= \frac{\vec{a}^T\vec{c}}{\vec{a}^T\vec{a}}\vec{a} \end{aligned} \tag{11.5}$$

11.2 Projecting to a 2D Subspace: Vector to Basis

Suppose that we know two data vectors, $\vec{a}_1$ and $\vec{a}_2$, that are linearly independent. They are a basis for some 2D subspace $\mathbb{V}$. Our problem becomes: what vector $\vec{p} \in \mathbb{V}$ is nearest to a new data vector $\vec{c}$?

The vector $\vec{p} \in \mathbb{V}$ can be written as a linear sum of the data vectors $\vec{a}_1$ and $\vec{a}_2$ because these vectors are a basis for $\mathbb{V}$. This entails that, for weights w_1 and w_2, we can write $\vec{p}$ as

$$\vec{p} \stackrel{\text{def}}{=} w_1\vec{a}_1 + w_2\vec{a}_2 \tag{11.6}$$

If we are given a new data vector $\vec{c}$, we want to find a vector $\vec{p}$ that is in the span of $\{\vec{a}_1, \vec{a}_2\}$ and that is "closest" to $\vec{c}$. This is illustrated in Figure 11.2.

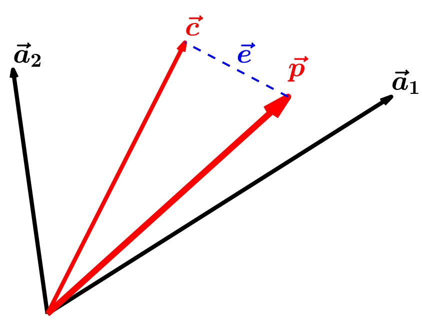

Figure 11.2: Vector $\vec{c}$ projects to $\vec{p}$, which is in a 2D subspace that is spanned by $\vec{a}_1$ and $\vec{a}_2$.

A numerical example may be helpful. Suppose that the data vectors are size 3 and that the known vectors span a plane that is described by the first two coordinates. Suppose also that the new data vector is $\vec{c}$. Examples of such vectors are

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{a}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \qquad \vec{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

We know that if we "drop" a perpendicular line in space, into the plane of the first two coordinates, that the result is "closest" to the new data vector $\vec{c}$ that we are studying[1]. In this example, we expect that the entries of $\vec{p}$ will be

$$\vec{p} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

We cannot, in general, deal with the vectors $\vec{a}_1$ and $\vec{a}_2$ separately. If we tried projecting $\vec{c}$ into the 1D subspace spanned by $\vec{a}_1$, and also tried projecting $\vec{c}$ into the 1D subspace spanned by $\vec{a}_2$, and combined the result, we would *not* get the expected projection. For the numerical examples above, the projections into the separate 1D subspaces are

$$\vec{p}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{p}_2 = \begin{bmatrix} 0.6 \\ 1.2 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \vec{p}_1 + \vec{p}_2 = \begin{bmatrix} 1.6 \\ 1.2 \\ 0 \end{bmatrix} \neq \vec{p}$$

The correct method is to deduce the "closest" vector $\vec{p}$ from geometry. We can begin by writing Equation 11.6 in matrix-vector form. Combining the data vectors $\vec{a}_1$ and $\vec{a}_2$ into a data matrix A, we can write the vector $\vec{p}$ as

$$\vec{p} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = A\vec{w} \tag{11.7}$$

We require the difference vector $\vec{e} = \vec{c} - \vec{p}$ to be orthogonal to the 2D subspace $\mathbb{V}$ that has $\vec{a}_1$ and $\vec{a}_2$ as a basis. If $\vec{e} \perp \mathbb{V}$, then $\vec{e}$ is orthogonal to the basis vectors of $\mathbb{V}$. From this we can deduce that

$$\begin{matrix} \vec{a}_1 \perp \vec{e} \\ \vec{a}_2 \perp \vec{e} \end{matrix} \quad \Rightarrow \quad \begin{matrix} \vec{a}_1^T \vec{e} = 0 \\ \vec{a}_2^T \vec{e} = 0 \end{matrix} \quad \Rightarrow \quad \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \end{bmatrix} \vec{e} = \vec{0} \quad \Rightarrow \quad A^T \vec{e} = \vec{0} \tag{11.8}$$

Using a matrix-vector form of the reasoning that we used for Equation 11.4, we find that

$$\begin{array}{rrcl} & A^T\vec{e} & = & \vec{0} \\ \Rightarrow & A^T(\vec{c} - \vec{p}) & = & \vec{0} \\ \Rightarrow & A^T\vec{c} - A^T\vec{p} & = & \vec{0} \\ \Rightarrow & A^T\vec{p} & = & A^T\vec{c} \\ \Rightarrow & A^T[A\vec{w}] & = & A^T\vec{c} \\ \Rightarrow & [A^TA]\vec{w} & = & A^T\vec{c} \end{array} \tag{11.9}$$

[1] In graphics this is called the *orthographic projection* of $\vec{c}$ into the plane

Equation 11.9 is called the *normal equation* for the data matrix A and the data vector $\vec{c}$. This equation describes how to find the weight vector $\vec{w}$ that projects the data vector $\vec{c}$ into the subspace $\mathbb{V}$ that has the columns of A as a basis.

By using the SVD of the matrix A, we know that if the columns of A are linearly independent then $[A^T A]$ is a symmetric positive definite matrix. This implies that Equation 11.9 has a single solution and the inverse $[A^T A]^{-1}$ exists. We can express the solution as a computation

$$\vec{w} = [A^T A]^{-1} A^T \vec{c} \tag{11.10}$$

To generalize Equation 11.5 to a higher-dimension subspace, we can abbreviate the process of projection as a linear transformation. The higher-dimension data vector $\vec{c}$ is projected to a lower-dimension subspace by a matrix P. From the expression for $\vec{p}$ in Equation 11.7, and the solution of the weight vector $\vec{w}$ in Equation 11.10, we can formulate the projection as

$$\begin{aligned} \vec{p} &= A\vec{w} \\ &= A[A^T A]^{-1} A^T \vec{c} \\ &= P\vec{c} \\ \Rightarrow \qquad P &= A[A^T A]^{-1} A^T \end{aligned} \tag{11.11}$$

Equation 11.11 describes how the projection works, from any vector of size m – that is, from any $\vec{c} \in \mathbb{R}^m$ – to any n-D subspace that has the columns of A as a basis.

