Projection matrices

We discussed projection matrices briefly when we discussed orthogonal projection. In particular, we discussed the following theorem.

Theorem. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$. Form the $n \times k$ matrix

$$\mathbf{U} = \left[\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \quad \ldots \quad & \mathbf{u}_k \end{array} \right].$$

Then $\text{proj}_W \mathbf{v} = \mathbf{U}\mathbf{U}^T\mathbf{v}$.

The matrix $\mathbf{U}\mathbf{U}^T$ is called the *projection matrix* for the subspace W. It does not depend on the choice of orthonormal basis.

What if we do not start with an orthonormal basis of W?

Theorem. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ be any basis for a subspace W of $\mathbb{R}^n$. Form the $n \times k$ matrix

$$\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \quad \ldots \quad & \mathbf{a}_k \end{array} \right].$$

Then the projection matrix for W is $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

A proof of this fact is posted on the course web site.

Note that any projection matrix $\mathbf{P}$ satisfies the two properties

1. $\mathbf{P}^2 = \mathbf{P}$, and
2. $\mathbf{P}$ is symmetric.

It is also true that any matrix that satisfies these two properties is the projection matrix for some subspace of $\mathbb{R}^n$.

Least squares approximation

Suppose we have data points $(x_1, y_1),\ (x_2, y_2), \dots,\ (x_n, y_n)$ and we want to "fit" them to a line.

How do we find the equation $y = mx + b$ of the line?

Form the matrices

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \text{and} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Note that we would have

$$\mathbf{Y} = \mathbf{X} \begin{bmatrix} b \\ m \end{bmatrix}$$

if all of the points were on the line $y = mx + b$.

Note also that we get the entire column space of $\mathbf{X}$ when we consider all vectors of the form

$$\mathbf{X} \begin{bmatrix} b \\ m \end{bmatrix}$$

where b and m are arbitrary.

Least-squares lines

We want to find m and b such that the quantity

$$||\mathbf{Y} - \mathbf{X}\begin{bmatrix} b \\ m \end{bmatrix}||$$

is as small as possible. Consequently,

$$\mathbf{X}\begin{bmatrix} b \\ m \end{bmatrix}$$

should be the projection of $\mathbf{Y}$ onto the column space of $\mathbf{X}$. Using the formula given earlier for the projection matrix, we have

$$\mathbf{X}\begin{bmatrix} b \\ m \end{bmatrix} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

Although $\mathbf{X}$ is not necessarily invertible, it has rank 2. Therefore, we can cancel $\mathbf{X}$ from the left on both sides. We obtain

$$\begin{bmatrix} b \\ m \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

Example. Suppose that a company had profits of \$500,000 in year 1, \$1,000,000 in year 2, and \$2,000,000 in year 5. Model its profits with a least-squares linear model.