DSC 40A Class Notes

Lecture 13
Friday, February 8, 2019

Our goal is to decompose $\vec{y}$ to write it as

$$\vec{y} = \vec{h} + \vec{e}$$

where $\vec{h} \in Col(X)$ and $e \perp Col(X)$.The following theorem says there is a unique way to decompose any vector $\vec{y}$ in this way.

Theorem 0.1 (Orthogonal Decomposition Theorem) *Each $\vec{y}$ in $\mathbb{R}^n$ can be written uniquely in the form $\vec{y} = \vec{h} + \vec{e}$ where $\vec{h} \in Col(X)$ and $\vec{e} \perp Col(X)$. This vector $\vec{h}$ is called the* orthogonal projection *of $\vec{y}$ onto the subspace $Col(X)$.*

Note that the error vector $\vec{e}$ must equal $\vec{y} - \vec{h}$.

Moreover, we can show that the orthogonal projection is actually the closest vector to $\vec{y}$ in $Col(X)$, and therefore the right choice of $\vec{h}$ to minimize our loss function. The following theorem is crucial; it says that our choice of $\vec{h}$ is optimal.

Theorem 0.2 (Best Approximation Theorem) *Let $\vec{h}$ be the orthogonal projection of $\vec{y}$ onto $Col(X)$. Then*

$$||\vec{y} - \vec{h}|| < ||\vec{y} - \vec{u}||$$

for all $\vec{u} \neq \vec{h}$ in $Col(X)$.

Proof Let $\vec{u} \neq \vec{h}$ in $Col(X)$. Then $\vec{h} - \vec{u}$ must also be in $Col(X)$. Explain why this is true.

col(x) is a subspace and when you add or subtract 2 vectors in a subspace, result is in subspace

The orthogonal decomposition theorem says we can write $\vec{y} = \vec{h} + \vec{e}$ with $\vec{h} \in Col(X)$ and $\vec{e} = (\vec{y} - \vec{h}) \perp Col(X)$. Since $\vec{h} - \vec{u}$ is in $Col(X)$, the definition of orthogonality to a subspace says that $\vec{e} = (\vec{y} - \vec{h}) \perp (\vec{h} - \vec{u})$. Then applying pythagorean theorem to the two orgthogonal vectors $\vec{y} - \vec{h}$

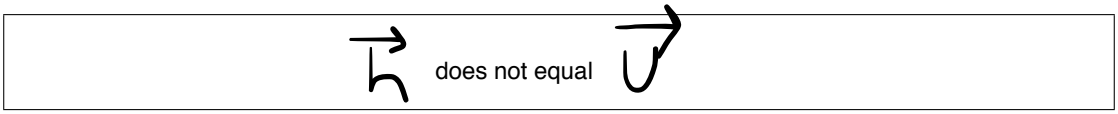
and $\vec{h} - \vec{u}$ gives

$$||(\vec{y} - \vec{h}) + (\vec{h} - \vec{u})||^2 = ||\vec{y} - \vec{h}||^2 + ||\vec{h} - \vec{u}||^2$$

or equivalently, from simplifying the left hand side,

$$||\vec{y} - \vec{u}||^2 = ||\vec{y} - \vec{h}||^2 + ||\vec{h} - \vec{u}||^2.$$

We know that $||\vec{h} - \vec{u}||^2$ is strictly positive, and not zero because



This means

$$||\vec{y} - \vec{h}||^2 < ||\vec{y} - \vec{u}||^2$$

or equivalently,

$$||\vec{y} - \vec{h}|| < ||\vec{y} - \vec{u}||$$

Since this argument holds for an arbitrary vector $\vec{u} \neq \vec{h}$ in $Col(X)$, we know that $||\vec{y} - \vec{h}|| < ||\vec{y} - \vec{u}||$ for all $\vec{u} \neq \vec{h}$ in $Col(X)$.

The Best Approximation Theorem says that the orthogonal projection of $\vec{y}$ onto $Col(X)$ is actually the closest possible vector to $\vec{y}$ that is in the column space of X. This says that the unique decomposition of $\vec{y}$ into $\vec{h}$ and $\vec{e}$ as guaranteed by the Orthogonal Decomposition Theorem coincides with the goal of minimizing our mean square error loss function.

The Orthogonal Decomposition guarantees that it is possible to decompose $\vec{y}$ into $\vec{h}$ and $\vec{e}$, but it does not specify how to do so. We can find $\vec{h}$ and $\vec{e}$ by writing

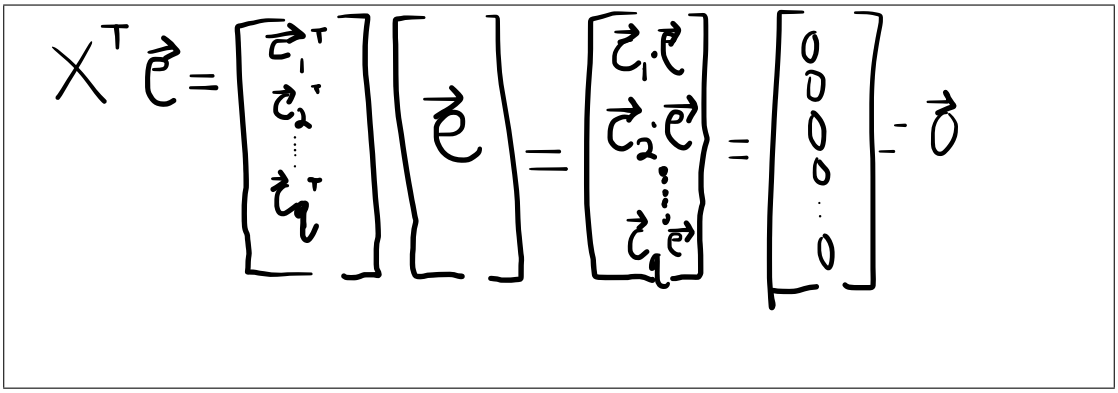
$$\vec{y} = \vec{h} + \vec{e} \tag{1}$$

$$\vec{y} = X\vec{b} + \vec{e} \tag{2}$$

$$X^T\vec{y} = X^TX\vec{b} + X^T\vec{e} \tag{3}$$

$$X^T\vec{y} = X^TX\vec{b} \tag{4}$$

To get from equation (2) to equation (3), we multiply on the left by X^T, the transpose of the design matrix X. To get from equation (3) to equation (4), we replace $X^T\vec{e}$ with $\vec{0}$. If $\vec{e} \perp Col(X)$ as guaranteed by the Orthogonal Decomposition Theorem, why is $X^T\vec{e} = \vec{0}$?



Equation (4) defines a system of linear equations, which are called the *normal equations*. We can solve the normal equation for $\vec{b}$ using methods from linear algebra to solve a system of equations. If the system has just one solution, we'll be able to invert the matrix X^TX and compute $\vec{b}$ as

$$\vec{b} = (X^TX)^{-1}X^T\vec{y},$$

but this is not always possible since not every matrix is invertible.

The components of $\vec{b}$ define the parameters of the best-fitting linear function $h(x) = b_0 + b_1x$. If needed, we could find the vector $\vec{h}$, the best approximation to $\vec{y}$ in $Col(X)$, by calculating $\vec{h} = X\vec{b}$, and we could find the error vector $\vec{e}$ by calculating $\vec{e} = \vec{y} - \vec{h}$.

Example: Find the linear function that best approximates the data $(2, 1)$, $(5, 2)$, $(7, 3)$, $(8, 3)$, and calculate the associated mean square error.

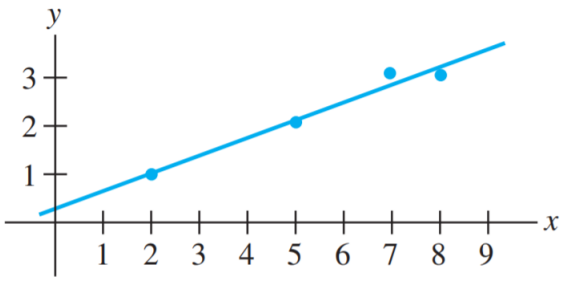
We have $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$ and $X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}$.

We can calculate $X^TX = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$ and $X^Ty = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, so the normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

Now, we just need to solve this linear system. The left-hand side matrix can be inverted to find $\vec{b} = \begin{bmatrix} \frac{2}{7} \\ \\ \frac{5}{14} \end{bmatrix}$, which means $\vec{h} = \begin{bmatrix} 1 \\ \frac{27}{14} \\ \frac{37}{14} \\ \frac{49}{14} \end{bmatrix}$ and $\vec{e} = \begin{bmatrix} 0 \\ \frac{1}{196} \\ \frac{9}{196} \\ \frac{1}{49} \end{bmatrix}$.

The best fit line is therefore $y = \frac{2}{7} + \frac{5}{14}x$, as shown in the picture below.



Calculate the mean square error associated with this line.

$(2,1)$ $(5,2)$ $(7,3)$ $(8,3)$

$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}$ $\vec{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$ $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$ $X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 37 \end{bmatrix}$

$\begin{bmatrix} 9 \\ 37 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$ $\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 2\frac{1}{7} \\ 5\frac{1}{14} \end{bmatrix}$

$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$

$\frac{5}{14} - \frac{4}{14} = \frac{1}{14}$

$\frac{1}{14}$

It is also useful to see where the normal equations come from in a particular example.We want to be able to write $\vec{y} = \vec{h} + \vec{e}$ where $\vec{h} \in Col(X)$ and $e \perp Col(X)$. Since $\vec{h} \in Col(X)$, this means $\vec{h} = X\vec{b}$ for some vector $\vec{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$. Since $\vec{e} \perp Col(X)$, this means

$$\vec{e} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 0 \text{ and } \vec{e} \cdot \begin{bmatrix} 2 \\ 5 \\ 7 \\ 8 \end{bmatrix} = 0.$$

We can write this equivlently as

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \vec{e} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $\vec{e} = \vec{y} - \vec{h} = \vec{y} - X\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$, substituting this expression into the

equation above gives

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \left(\begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This simplifies to

$$\begin{bmatrix} 9 \\ 57 \end{bmatrix} - \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

or

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

This is the same linear system that we got from using the normal equations. Either approach can be used to get to the same linear system.