# Response to FtYou

answered by mixedmath
David Lowry-Duda
Brown University Mathematics
http://davidlowryduda.com

FtYou writes

> Hello everyone ! There is a concept I have a hard time getting my head wrap around. If you have a Vector Space V and a subspace W, I understand that you can find the least square vector approximation from any vector in V to a vector in W. And this correspond to the projection of V to the subspace W. Now , for data fitting ... Let's suppose you have a bunch of points (xi, yi) where you want to fit a set a regressors so you can approximate yi by a linear combination of the regressors lets say ( 1, x, x2 ... ). What Vector space are we talking about ? If we consider the Vector space of function R -¿ R, in what subspace are we trying to map these vectors ? I have a hard time merging these two concepts of projecting to a vector space and fitting the data. In the latter case what vector are we using ? The functions ? If so I understand the choice of regressors ( which constitute a basis for the vector space ) But what's the role of the (xi,yi) ?
>
> I want to point out that I understand completely how to build the matrices to get Y = AX and solving using least square approx. What I miss is the big picture. The linear algebra picture. Thanks for any help !

We'll go over this by closely examining and understanding an example. Suppose we have the data points $(x_i, y_i)$

$$\begin{cases} (x_1, y_1) = (-1, 8) \\ (x_2, y_2) = (0, 8) \\ (x_3, y_3) = (1, 4) \\ (x_4, y_4) = (2, 16) \end{cases},$$

and we have decided to try to find the best fitting quadratic function. What do we mean by best-fitting? We mean that we want the one that approximates these data points the best. What exactly does that mean? We'll see that before the end of this note - but in linear algebra terms, we are projecting on to some sort of vector space - we claim that projection is the "best-fit" possible.

So what do we do? A generic quadratic function is $f(t) = a + bt + ct^2$. Intuitively, we apply what we know. Then the points above become

$$\begin{cases} f(-1) = a - b + c = 8 \\ f(0) = a = 8 \\ f(1) = a + b + c = 4 \\ f(2) = a + 2b + 4c = 16 \end{cases},$$

and we want to find the best $[abc]$ we can that "solves" this. Of course, this is a matrix equation:

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix}.$$

And so you see how the algorithm would complete this. But now let's get down the "linear algebra picture," as you say.

We know that quadratic polynomials $f(t) = a + bt + ct^2$ are a three dimensional vector space (which I denote by $P_2$) spanned by $1, t, t^2$. We know we have four data points, so we will define a linear transformation $A$ to be the transformation taking a quadratic polynomial $f$ to $\mathbb{R}^4$ by evaluating $f$ at $-1, 0, 1, 2$ (i.e. the $x_i$). In other words,

$$A : P_2 \longrightarrow \mathbb{R}^4$$

where

$$A(f) = \begin{pmatrix} f(-1) \\ f(0) \\ f(1) \\ f(2) \end{pmatrix}.$$

We interpret $f$ as being given by three coordinates, $a, b, c \in \mathbb{R}^3$, so we can think of $A$ as a linear transformation from $\mathbb{R}^3 \longrightarrow \mathbb{R}^4$. In fact, $A$ is nothing more than the matrix we wrote above.

Then a solution to

$$A^T A \begin{pmatrix} a \\ b \\ c \end{pmatrix} = A^T \begin{pmatrix} 8 \\ 8 \\ 4 \\ 16 \end{pmatrix}$$

is the projection of the space of quadratic polynomials on $\mathbb{R}^4$ (which in this case is the space of evaluations of quadratic polynomials at four different points). If $f^*$ is the found projection, and I denote the $y_i$ coordinate vector as $y^*$, then this projection minimizes

$$||y^* - Af^*||^2 = (y_1 - f^*(x_1))^2 + \ldots + (y_4 - f^*(x_4))^2,$$

and it is in this sense that we mean we have the "best-fit." (This is roughly interpreted as the distances between the $y_i$ and $f^*(x_i)$ are minimized; really, it's the sum of the squares of the distances - hence "Least-Squares").

So in short: $A$ is a matrix evaluating quadratic polynomials at different points. The columns vectors correspond to a basis for the space of quadratic polynomials, $1, t, t^2$. The codomain is $\mathbb{R}^4$, coming from the evaluation of the input polynomial at the four different $x_i$. The projection of the set of quadratic polynomials onto their evaluation space minimizes the sum of the squares of the distances between $f(x_i)$ and $y_i$.

Does that make sense?