This material will not be on the midterm or the final! I indicate it here as an application of the some of the linear algebra we've seen this week.

Suppose you're a researcher studying the effects of the dosages of two medicines on the lifespans of mice. You might collect the following data on 100 mice (units omitted):

	Dose of Medicine 1	Dose of Medicine 2	Lifespan
Mouse 1	250	90	2
Mouse 2	200	100	2.5
Mouse 3	150	150	1.5
⋮	⋮	⋮	⋮
Mouse 100	110	170	2.7

Suppose you think that, approximately

$$\text{Lifespan} \approx a_1\,(\text{Dose of Medicine 1}) + a_2\,(\text{Dose of Medicine 2})$$

for some numbers a_1 and a_2. That is, suppose you think that Lifespan is a linear function of the doses of the two medicines. Then you would like to estimate from your data the coefficients a_1 and a_2. Let v_1 be the column vector for the doses of medicine 1:

$$v_1 = \begin{pmatrix} 250 \\ 200 \\ 150 \\ \vdots \\ 110 \end{pmatrix}$$

let v_2 be the column vector for the doses of medicine 2:

$$v_2 = \begin{pmatrix} 90 \\ 100 \\ 150 \\ \vdots \\ 170 \end{pmatrix}$$

and let w be the column vector of lifespans:

$$w = \begin{pmatrix} 2 \\ 2.5 \\ 1.5 \\ \vdots \\ 2.7 \end{pmatrix}.$$

If it were exactly true that

$$\text{Lifespan} = a_1\,(\text{Dose of Medicine 1}) + a_2\,(\text{Dose of Medicine 2})$$

for some numbers a_1 and a_2, then it would be true that

$$w = a_1 v_1 + a_2 w_2$$

so w would be in the span of v_1 and v_2. In general, w won't be in the span of v_1 and v_2. However, let $\widetilde{w}$ be the orthogonal projection of w to the span of v_1 and v_2. Then $\widetilde{w}$ is the closest point to w in the span in the span of v_1 and v_2. Therefore $\widetilde{w} = a_1 v_1 + a_2 v_2$ for some a_1 and a_2 and $\widetilde{w}$ is the closest such vector. If the assumption that

$$\text{Lifespan} \approx a_1\,(\text{Dose of Medicine 1}) + a_2\,(\text{Dose of Medicine 2})$$

is good, then w will be close to $\widetilde{w}$ and

$$w \approx a_1 v_1 + a_2 v_2$$

Therefore to find a_1 and a_2, the researcher should solve the equation $\widetilde{w} = a_1 v_1 + a_2 v_2$ for a_1 and a_2 from the known quantities v_1, v_2, and $\widetilde{w}$.

Note that v_1, v_2 and w are vectors in $\mathbb{R}^{100}$. If 1000 mice were measured, they would be vectors in $\mathbb{R}^{1000}$ instead. Also note that v_1 and v_2 each correspond to one of the two independent variables. If there were fifteen independent variables instead of there would be $v_1, \ldots, v_{15}$ instead of v_1, v_2.

This is usually notated in a different but helpful way. Let X be a 100×2 matrix whose columns form the data of the independent variables, i.e.,

$$X = \begin{pmatrix} 250 & 90 \\ 200 & 100 \\ 150 & 150 \\ \vdots & \vdots \\ 110 & 170 \end{pmatrix}$$

Let y be a column matrix whose entries are the dependent variable measurements

$$y = \begin{pmatrix} 2 \\ 2.5 \\ 1.5 \\ \vdots \\ 2.7 \end{pmatrix}$$

(so y is what was called w before and X is the matrix whose columns were called v_1 and v_2). Then the approximation $w \approx a_1 v_1 + a_2 v_2$ can be rewritten

$$y \approx X\beta$$

where

$$\beta = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

You want to find β such that $X\beta$ is the orthogonal projection of y to the span of the columns of X. Therefore you want to find β that minimizes the length of the vector $\|y - X\beta\|$ as β ranges over all 2×1 matrices. This is the same as finding β that minimizes the length squared: $\|y - X\beta\|^2$. Some simple multivariable calculus shows that, if $X^T X$ is invertible (and, in practice, it usually is), then

$$\beta = (X^T X)^{-1} X^T y.$$

is the unique such minimizer.