Application: multiple linear regression

> In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables.
>
> The case of one explanatory variable is called simple linear regression.
>
> For more than one explanatory variable, the process is called multiple linear regression.
>
> `http://en.wikipedia.org/wiki/Linear_regression`

The experimental data might be of the form (x_i, y_i, z_i), where now the dependent variable z_i depends on two explanatory variables x_i, y_i (instead of just x_i).

Example 48. Set up a linear system to find values for the parameters a, b, c such that $z = a + bx + cy$ best fits some given points $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots$

Solution. The equations $a + bx_i + cy_i = z_i$ translate into the system:

$$\underbrace{\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } A} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector } \boldsymbol{z}}}$$

Of course, this is usually inconsistent. To find the best possible a, b, c we compute a least squares solution by solving $A^T A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = A^T \boldsymbol{z}$.

More on orthogonality

Projection matrices

Review. We can compute the orthogonal projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto W as follows:

- Write $W = \text{col}(A)$, where the columns of A are a basis of W.
  Then, $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ where $\hat{\boldsymbol{x}}$ is the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ (i.e. $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$).

Assuming $A^T A$ is invertible (which, as we will see next class, is automatically the case if the columns of A are independent), we have $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$ and hence:

(projection matrix) The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\text{col}(A)$ is $\qquad$ (assuming cols of A are independent)

$$\hat{\boldsymbol{b}} = \underbrace{A(A^T A)^{-1} A^T}_{P} \boldsymbol{b}.$$

The matrix $P = A(A^T A)^{-1} A^T$ is the **projection matrix** for projecting onto $\text{col}(A)$.

Example 49.

(a) What is the matrix P for projecting onto $W = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$?

(b) Using the projection matrix, project $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$ onto $W = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Solution.

(a) Choosing $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, the projection matrix P is $A(A^TA)^{-1}A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$

$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 2 & -4 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Comment. We can choose A in any way such that its columns are a basis for W. The final projection matrix will always be the same.

(b) The projection is $\frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix}$.

Check. The error $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$ is indeed orthogonal to W.

Projecting onto 1-dimensional spaces

When we project onto a 1-dimensional space $\text{span}\{w\}$, we usually just say that we are projecting onto w.

> The (orthogonal) projection of v onto w is $\dfrac{w \cdot v}{\|w\|^2} w$.

Why? Replace b with v and A with w in our general projection matrix formula to get $w(w^Tw)^{-1}w^Tv$, which equals $\dfrac{w \cdot v}{\|w\|^2} w$ (note that $w^Tv = w \cdot v$ and $w^Tw = \|w\|^2$ are scalars).

Comment. If you have taken Calculus 3, you have seen that formula before. Most likely, you were deriving it using angles at that time. Namely, the dot product has the following connection to angles:

> $v \cdot w = \|v\| \, \|w\| \cos\theta$ where $\theta \in [0, \pi]$ is the angle between v and w

Why? You can derive this by repeating what we did, right after Definition 19 to show that v and w are orthogonal if and only if $v \cdot w = 0$. Just replace Pythagoras with the law of cosines ($c^2 = a^2 + b^2 - 2ab \cos\theta$ holds in any triangle!).

Two obvious cases. Observe that the cases $\theta = 0$ and $\theta = 90°$ are clearly true.

We will not discuss angles much further in this class. Just in case it is helpful, here is the typical argument given in Calculus 3 to determine the projection $\text{proj}_w v$ of v onto w:

From the sketch, we see that "error" $= v - \text{proj}_w v$ and that this error is orthogonal to w.

Basic trigonometry tells us that the length of $\text{proj}_w v$ is $\|v\| \cos\theta$. Hence:

$$\text{proj}_w v = \underbrace{\|v\| \cos\theta}_{\text{length}} \underbrace{\frac{w}{\|w\|}}_{\text{direction}}$$

$$= \frac{\|v\| \, \|w\| \cos\theta}{\|w\|} \frac{w}{\|w\|} = \left(\frac{v \cdot w}{\|w\|^2} \right) w$$