The geometry of least squares

We can think of a vector as a point in space, where the elements of the vector are the coordinates of the point. Consider for example, the following vector's:

$$t = (-4, 0), \quad u = (5, 0), \quad v = (4, 4), \quad w = (-2, 2)$$

Each of these vectors has two elements. If we regard, for example, the first element of the vector **v** as the x-coordinate and the second element of **v** as the y-coordinate, each vector can be represented by a point in a two-dimensional space as shown in the picture below.

This idea can be extended to factors with three or more elements, but we need a space of three or more dimensions to do so. Our inability to visualize a multidimensional space, however, should not prevent us from thinking of any n-element vector as a point in an n-dimensional space

A vector can also be viewed geometric way as they directed line segment with an arrow starting from the origin and ending at the point representing the vector in the space. Draw this geometric representation in the plot below.

Again, this way of viewing vectors can be extended to vectors with three or more elements, as we can always imagine a line segment starting at the origin of an n-dimensional space and ending at the point representing the vector in the space.

This geometric representation of vectors raises 2 questions:
What are the length and direction of the line segment representing a vector?
What is the angle between two line segments representing two vectors?

We now have three ways of viewing a vector: algebraically, graphically, and geometrically. Let's measure the length of the line segment representing the vector $\mathbf{v}$. Using Pythagorean theorem, the length of the vector $\mathbf{v}$, that is, the line from the origin to the point $\mathbf{v}$ in the figure above is:

You can verify that the lengths of the vectors $\mathbf{w}$ and $\mathbf{t}$ are $2\sqrt{2}$ and 4, respectively.

More generally, for any n-dimensional vector $\mathbf{v} = (v_1, v_2, ..., v_n)$, the length of $\mathbf{v}$, denoted by $||\mathbf{v}||$ is defined as

$$||v|| = \sqrt{v'v} =$$

The length the vector is also referred to as the norm of the vector. Note that the length of the geometric vector can be interpreted as the magnitude of the corresponding algebraic vector. Thus the larger the elements of the vector in absolute value, the greater its length and the larger the magnitude of the vector.

Now let's measure the angle between any to geometric vectors. Consider the angle between the two vectors in the diagram below.

The numerator is the inner product of the two vectors. Thus, the cosine of the angle between two vectors is the ratio of their inner product and the product of their lengths (norms). For example, let's measure the angle between the vectors u and v from above.

$$
\begin{aligned}
u'v &= 5 * 4 + 0 * 4 = 20 \\
||u|| &= 5 \\
||v|| &= 4\sqrt{2} \\
\text{And} \quad \text{hence} \quad \cos(\theta) &= \frac{u'v}{||u||||v||} = \frac{1}{\sqrt{2}}
\end{aligned}
$$

The angle between two vectors is directly related to the concepts of linear dependence and independence. Two vectors **u** and **v** are linearly dependent if and only if one vector can be written as the scalar multiple of the other. When this happens, the cosine of the angle between these two vectors is either +1 or -1. This implies that the angle is either 0 or 180 degrees. We may then conclude that went to vectors are linearly dependent, the angle between them is either 0 or 180 degrees, which means that the two line segments representing the vectors have the same (or opposite) direction. The converse of the statement is also true.

Two vectors **u** and **v** of the same order are perpendicular if and only if $u'v = 0$. We also say that these two vectors are orthogonal. The converse is also true, that is, when two vectors have an inner product of 0 then they are orthogonal. However if two vectors are linearly independent they may or may not be orthogonal.

Now take a look at what happens graphically only after subtract two vectors, or when we multiply vector by a scalar. Suppose you wish to add the vectors **v**. and **w**. Algebraically, we have

Draw the geomtric picture below. Notice that the line segment representing the sum of the two vectors is the diagonal of the parallelogram having the line segments representing the two vectors as adjacent to edges. This is known as the parallelogram rule, and is always true for the sum of any two vectors.

What about the difference between two vectors? The distance between the vectors **v**. and **w** is the same as the norm (length) of the vector (**v-w**). It is sometimes useful to consider the difference between two vectors **v** and **w** as the sum of **v** and **-w**.

Multiplying a vector by a scalar amounts to multiplying each of its elements by the scalar. Draw a picture to convince yourself of this.

A vector of length one is called a normal vector. Any non-zero vector can be normalized by dividing each of its elements by its norm. This transformation is called normalization.

The normalized versions of two vectors of the same direction are equal.

The elements of the normal vector can be interpreted as direction cosines. Simply stated, they are the cosine of the angles between the vector and each of the coordinate axes. Draw a picture to convince yourself of this.

Connection to Fitting by Least Squares

Consider observations $(x_1, y_1), \ldots, (x_n, y_n)$. Reexpress these observations as two n-dimensional vectors, $\mathbf{y} = (y_1, \ldots, y_n)$ and $\mathbf{x} = (x_1, \ldots, x_n)$. Let $\mathbf{1}$ denote the vector in n-dimensional space of all 1's, i.e. $\mathbf{1} = (1, 1, \ldots, 1)$.

The span of $\mathbf{x}$ is all those vectors of the form c$\mathbf{x}$, where c is a scalar. The span of $\mathbf{x}, \mathbf{1}$ is the collection of vectors that can be expressed as $a\mathbf{x} + b\mathbf{1}$.

Show the following

1. $\sum_{i=1}^{n} (y_i - c)^2 = || \qquad\qquad\qquad\qquad ||^2$
2. So minimizing the quantity above with respect to c is the same as finding the ______ in the linear span of ______
3. Show that $\mathbf{1}$ is orthogonal to $\mathbf{y} - \bar{y}\mathbf{1}$. Use a picture to verify this fact.

4. Use this fact to show that

$$||\mathbf{y} - c\mathbf{1}||^2 = ||\mathbf{y} - \bar{y}\mathbf{1}||^2 + ||(\bar{y} - c)\mathbf{1}||^2$$

5. Next establish that $\bar{y}$ is the minimizer of $\sum_{i=1}^{n}(y_i - c)^2$, and that $\bar{y}\mathbf{1}$ is the closest point in the linear span of $\mathbf{1}$ to $\mathbf{y}$.

6. Show that in general the closest vector to $\mathbf{y}$ in the linear span of a vector $\mathbf{v}$ is the projection

$$P_v\mathbf{y} = \frac{y'v}{||v||^2}\mathbf{v}$$

7. Now consider the lest squares fit from this geometric perspective

$$\sum_{i=1}^{n}[y_i - (a + bx_i)]^2 = || \qquad\qquad\qquad\qquad\qquad\qquad ||^2$$

Minimizing with respect to a and b is equivalent to projecting ______________ onto the linear span of ______________ and ______________, of finding the closest vector to ______________ in the linear span of ______________

8. Show that the linear span of $\mathbf{1}$ and $\mathbf{x}$ is the same as the linear span of $\mathbf{1}$ and $\mathbf{x} - \bar{x}\mathbf{1}$.

9. Explain why $P_{1,x}\mathbf{y} = P_1\mathbf{y} + P_{x-\bar{x}1}\mathbf{y}$.

10. Show that $P_1\mathbf{y} = \bar{y}\mathbf{1}$ and

$$P_{x-\bar{x}1}\mathbf{y} = \frac{y'(x - \bar{x}1)}{||x - \bar{x}1||^2}(x - \bar{x}1)$$

11. Show that $\hat{b}$ obtained from minimizing $\sum_{i=1}^{n}[y_i - (a + bx_i)]^2$ $\frac{y'(x-\bar{x}1)}{||x-\bar{x}1||^2}$

12. Reexpress $\hat{a}$ as well.

13. We now see that $\hat{\mathbf{y}}$ is the projection of ______________ onto the linear span of ______________. We also see that the residuals $\mathbf{e}$ are ______________ to $\hat{\mathbf{y}}$, i.e. $\mathbf{e} = \mathbf{y} - P_{1,x}\mathbf{y}$.