1: 2

Vector math

A vector is simply a collection of n numbers. You can think of a vector as coordinates in an n dimensional space. Since it is difficult to visualize an n dimensional space, lets restrict these examples to vectors with two elements ($n = 2$). If we have a vector $\mathbf{v}$ with two elements (1,2), then you can plot this in a two-dimensional x, y coordinate system. If we fix the vector at the origin (0,0) and plot v we can see that the vector has a direction and length (or magnitude). The magnitude of a vector is the square root of the sum of the squared elements $\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}} = \sqrt{\sum_i^n v_i^2}$

Multiplying vector by scalar

A vector:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

If we multiply a vector by scalar, we just extend its length

$$\Theta = 2, \mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\mathbf{u} = \Theta \mathbf{v} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

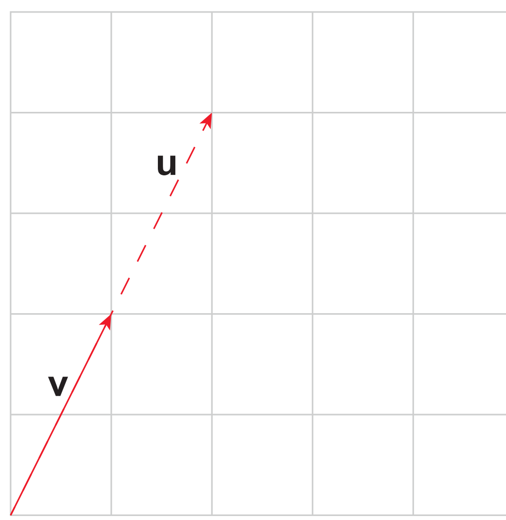

Figure 1. Multiplying vector by scalar.

Vector addition

The two vectors

must have the same number of elements. We just

add the corresponding elements in each vector.

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{p} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

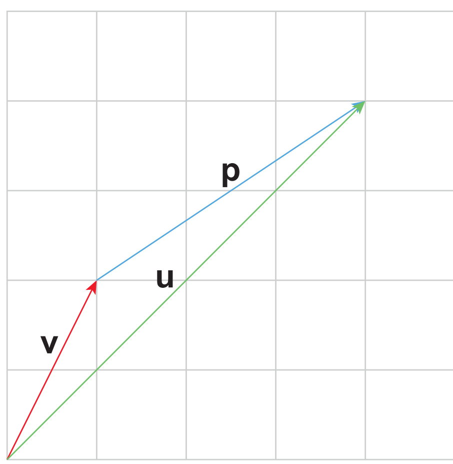

Figure 2. Vector addition

The linear model

If we are interested in studying the relationship between two variables (x and y) we can express this relationship in a linear model.

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

The linear model is just a collection of vectors. The equation above is for one element i in the x and y vectors. We can also express this as vectors

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Grouping the predictor vectors into a matrix

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

There's some things to clarify here. We can decompose $\mathbf{y}$ into a systematic component defined by the linear combination of predictors (e.g. $\beta_0 + \beta_1 x_i$) and a stochastic or random component (i.e. e_i). The error term/vector $\mathbf{e}$ is random because our experiment/sampling is the result of a random process. It is associated with a probability density function (pdf), specifically a Normal pdf with a mean of 0 and a variance of σ^2. The most likely value for e will be 0. If we add a fixed vector to a random variable, the result will be another random variable. Thus we can say that our vector of observations, $\mathbf{y}$, is just one realization of a random vector, $\mathbf{Y}$ and is $\mathbf{Y} \sim N(\beta_0 \mathbf{1} + \beta_1 \mathbf{x}, \sigma^2)$. We are taking the density of the error and shifting it by $\beta_0 \mathbf{1} + \beta_1 \mathbf{x}$. We are moving the peak (the mean), or location of the distribution.

Hopefully this example will provide some clarification. Suppose we have two vectors and each have two elements. We have a random vector, $\mathbf{e}$, and a fixed vector μ. $\boldsymbol{\mu}$ is the systematic component of the linear model was calculated as follows

$$\mu = \begin{pmatrix} 3 \\ 1 \end{pmatrix} = 0.5 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 1.25 \\ 0.25 \end{pmatrix}$$

.

So $\mu_i = \beta_0 + \beta_1 x_i$. The random component is $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I}_2)$. The $\mathbf{I}_n$ is a 2×2 matrix with 1 in the diagonal and 0's in all off diagonal elements (i.e. it is a 2×2 identity matrix). This coordinates for this vector is (0, 0). We can visualize them in a 2 dimensional space as:

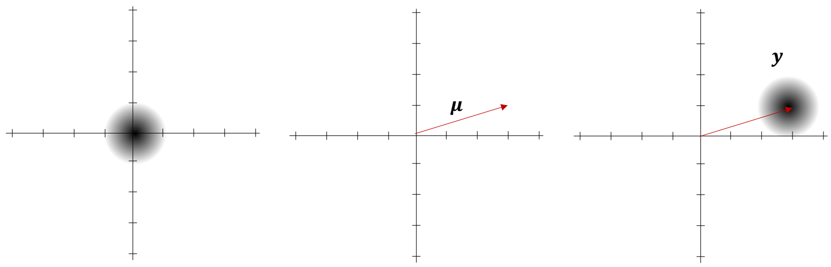

Figure 3. Visualizing y.

So our observed values for the response variable can exist anywhere within the point cloud on the right-most panel.

Vector space

We assume the predictor variables (and their associated coefficients) are fixed. They have a fixed endpoint. When we multiply the predictor variables by their coefficients it results in another fixed vector. It can be visualized as a line. All out predictor vectors exist in a 2D plane. The goal of ordinary least squares is to project the vector **y** into this 2D plane. This is where this term "projection" or where the projection matrix (also called the hat matrix) comes from. The dashed vector in the figure below is the projection of **y** into the predictor plane.

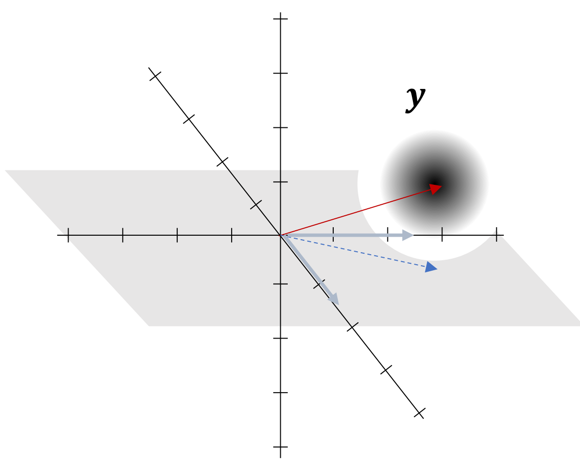

Figure 4. We are basically just extending the concepts in Figure 3 into 3 dimensions. In other words, rather than showing the mean vector $\boldsymbol{\mu}$ we are showing the individual vectors for the systematic component of the linear model. The blue dashed line is $\hat{\mathbf{y}}$

So in OLS we are trying to find the best-fit line (the values of $\hat{\mathbf{y}}$ that are closest to $\mathbf{y}$). From Figure 4, we can see that the best fit is directly below $\mathbf{y}$ (think of it as the shadow of $\mathbf{y}$ at high-noon). The distance from the point of $\mathbf{y}$ to the tip of $\hat{\mathbf{y}}$ is the norm of the vector of residuals.

The projection matrix is

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$$

You can see that is projects $\mathbf{y}$ onto the predictor space by looking how we get the fitted values

$$\hat{\mathbf{y}} = \boldsymbol{X\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$$

Partitioning the sum of squares

Moving bask to the two-dimensional example shown in Figure 3. So were combining all systematic effects into one vector $\hat{\mathbf{y}}$.

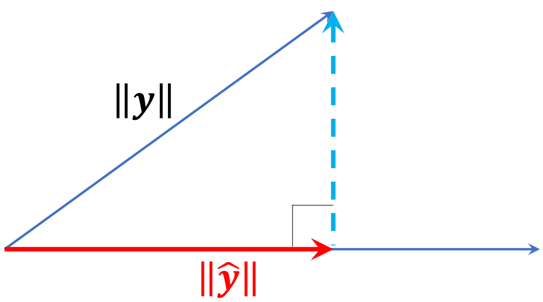

Figure 5. Partitioning the sum of squares. The blue dashed vector is the vector of residuals.

We know that the length of a vector is given by

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} v_i^2$$

Then we can show that with OLS we are partitioning the total sum of squares into orthogonal components: sum of squares for the regression model (systematic part of the linear model) and the residual sum of squares.

$$\begin{aligned}\|\mathbf{y}\|^2 &= \|\hat{\mathbf{e}}\|^2 + \|\hat{\mathbf{y}}\|^2 \\ \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} \hat{e}_i^2 + \sum_{i=1}^{n} \hat{y}_i^2\end{aligned}$$