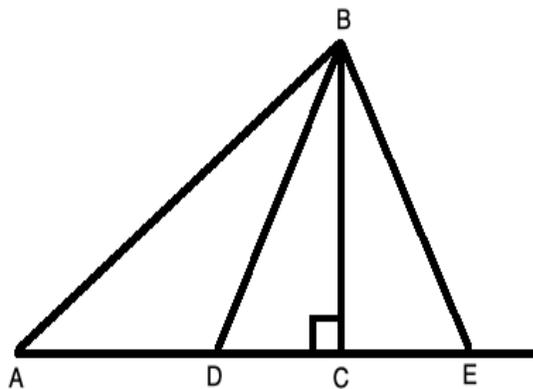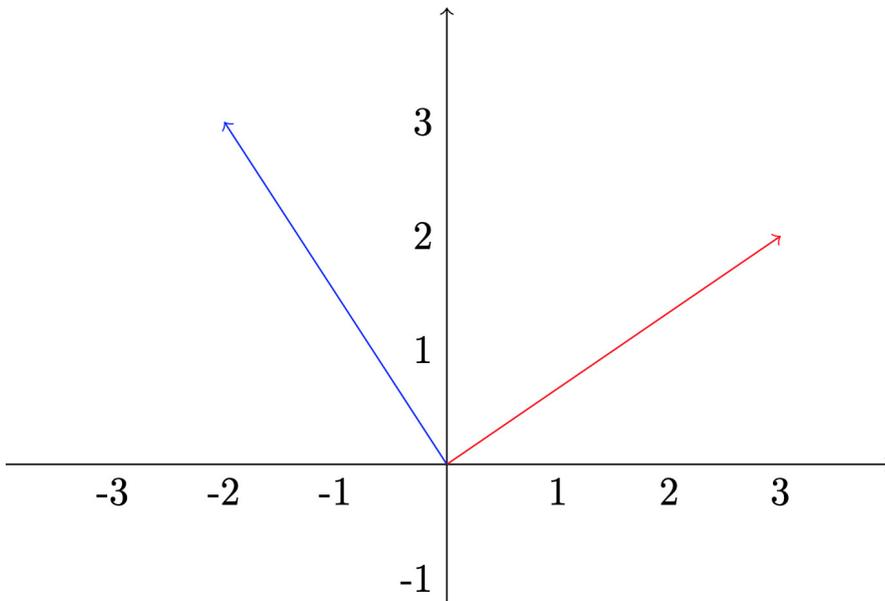# OLS Geometry

(Jing Li, Miami University)

1. This notes explains how to use geometric intuition to solve the least squares or line fitting problem and obtain the OLS estimator. It is motivated by the observation that some students struggle in calculus course, and therefore have a hard time understanding how to use calculus (i.e., taking derivative and setting it to zero) to find solution.

2. The least squares problem belongs to a broad problem of optimization. Without losing generality, this note focuses on minimization since maximization can be transformed to minimizing the negative of original objective.

3. In geometry class, we know how to solve the problem of <u>minimizing the distance</u> between point B and line AE, see the picture below. The solution, or the shortest distance, is line BC, which is <u>perpendicular</u> to line AE.

Orthogonal (Perpendicular) Condition For Minimizing Distance

It is clear that BC is shorter than BD and BE, and BC and AC are perpendicular (their angle is 90 degree). Notice how easy it is to grasp the basic idea of solving a minimization problem with geometry compared to calculus.

4. Line AB can represent a <u>vector</u> if we add direction (arrow) to it. In that case, line AC is <u>projection</u> of AB onto AE, and line BC is called <u>projection error</u>. Notice that AC is a portion of AE, or $AC = cAE$, where c denotes a multiplier or coefficient. Finding the shortest distance amounts to finding the optional multiplier c.

5. The key insight is that the distance is minimized if projection and projection error are perpendicular. We call that <u>orthogonal condition</u>.

6. Next, we switch from lines to numbers, and we wonder how to <u>generalize</u> the idea of two lines being perpendicular. The graph below, which displays the rectangular coordinate system (Cartesian plane), provides a hint

The red line represents the vector (3,2), and blue line represents (-2,3). We see the two lines are perpendicular!

7. They are perpendicular because

$$\texttt{product of x coordinates} + \texttt{product of y coordinates} = (3)(-2) + (2)(3) {\color{red}= 0} \quad (1)$$

This finding motivates the following definition — the <u>dot product</u> or inner product of two vectors (x_a, y_a) and (x_b, y_b) is defined as

$$a \cdot b \equiv x_a x_b + y_a y_b \qquad (\texttt{dot product}) \qquad (2)$$

So for two vectors, the dot product returns a scalar.

8. Dot product can be used to measure length or size of a vector. According to the Pythagorean Theorem, the Euclidean norm or length of the vector (x_a, y_a) is

$$||a||_2 = \sqrt{x_a^2 + y_a^2} = \sqrt{a \cdot a} \qquad (\texttt{Euclidean norm or length}) \qquad (3)$$

9. Dot product is also related to angle. Let θ_a be the angle between vector (x_a, y_a) and horizontal axis, and θ_b be the angle for (x_b, y_b). The angle between the two vectors is $\theta_b - \theta_a$. The trigonometry class teaches us that
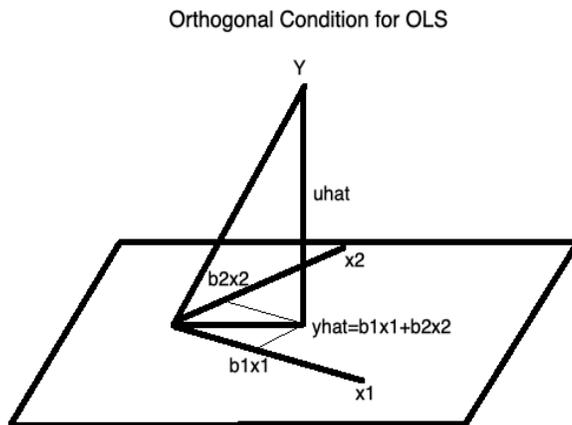
$$cos(\theta_b - \theta_a) = cos(\theta_b)cos(\theta_a) + sin(\theta_b)sin(\theta_a)$$
$$= \frac{x_b}{||b||_2}\frac{x_a}{||a||_2} + \frac{y_b}{||b||_2}\frac{y_a}{||a||_2} = \frac{x_a x_b + y_a y_b}{||a||_2||b||_2} = \frac{a \cdot b}{||a||_2||b||_2} \qquad (4)$$

Thus, dot product being zero implies that two vectors are perpendicular:

$$a \cdot b = 0 \Rightarrow cos(\theta_b - \theta_a) = 0 \Rightarrow \theta_b - \theta_a = 90° \Rightarrow (x_a, y_a) \text{ and } (x_b, y_b) \text{ are perpendicular} \qquad (5)$$

To sum up, the geometric idea of "two lines being perpendicular" is extended to "dot product being zero".

10. Now we are ready to use geometric intuition to solve the least squares problem, see the graph below

Orthogonal Condition for OLS

Note that

3

(a) Y is an n by 1 vector that consists of n observations of the outcome variable: $Y = (y_1, y_2, ... y_n)$. In the n-dimensional space (hard to imagine, I know) denoted by $\mathbb{R}^n$, it is represented by a point—y_1 is its first coordinate, y_2 is the second coordinate, and so on. The line connecting the point and origin is shown in the graph.

(b) Suppose there are two predictors $x1$ and $x2$, each being an n by 1 vector represented by a line. Those two x lines form a <u>column space</u>, which is a sub-space in $\mathbb{R}^n$. Any linear combination of the two vectors $b_1 x1 + b_2 x2$ lies in that subspace.

(c) Ordinary Least Squares (OLS) aim to find the best projection of Y onto that column space spanned by $x1$ and $x2$. That is equivalent to finding the best coefficients or multipliers b_1 and b_2. The projection is the best in the sense that the <u>distance</u> between Y and the column space is <u>minimized</u>.

(d) We call upon the intuition that projection error need to be perpendicular to column space in order for the distance to be the least. Furthermore, being perpendicular requires zero dot product. It follows that

```
projection error is perpendicular to column space ⇒
```
$$\text{projection error is perpendicular to } x_1 \text{ } and$$
$$\text{projection error is perpendicular to } x_2 \Rightarrow$$
$$\hat{u} \cdot x_1 = 0 \text{ } and \text{ } \hat{u} \cdot x_2 = 0 \Rightarrow$$
$$\sum_{i=1}^{n} x_{1i} \hat{u}_i = 0 \text{ } and \text{ } \sum_{i=1}^{n} x_{2i} \hat{u}_i = 0 \quad \text{(Orthogonal Conditions of OLS)} \quad (6)$$

where $\hat{u} = y - \hat{y}$ is the residual (projection error), and $\hat{y} = b_1 x_1 + b_2 x_2$ is fitted value (projection). Notice that in $\mathbb{R}^n$ space, the dot product is sum of products of n coordinates.

(e) Those two orthogonal conditions are the same as the two first order conditions derived by using calculus to minimize residual sum squares

$$\frac{\partial \sum_i (y_i - \hat{y}_i)^2}{\partial b_1} = 0 \Rightarrow \frac{\partial \sum_i (y_i - b_1 x_{1i} - b_2 x_{2i})^2}{\partial b_1} = 0 \Rightarrow \sum_i x_{1i} \hat{u}_i = 0 \quad (7)$$

$$\frac{\partial \sum_i (y_i - \hat{y}_i)^2}{\partial b_2} = 0 \Rightarrow \frac{\partial \sum_i (y_i - b_1 x_{1i} - b_2 x_{2i})^2}{\partial b_2} = 0 \Rightarrow \sum_i x_{2i} \hat{u}_i = 0 \quad (8)$$

4

As a result, using geometry leads to the same OLS estimators as using calculus.

11. We can use geometric approach to prove many difficult results. For instance, according to the previous result

$$cos(\theta_b - \theta_a) = \frac{x_a x_b + y_a y_b}{||a||_2 ||b||_2}$$

and because $-1 \leq cos \leq 1$, it follows that

$$(x_a x_b + y_a y_b)^2 \leq (x_a^2 + y_a^2)(x_b^2 + y_b^2) \quad \text{(Cauchy Schwarz Inequality)} \quad (9)$$

Extend the Cauchy–Schwarz inequality to $\mathbb{R}^n$ space, we can prove that the correlation coefficient ρ is bounded by -1 and 1

$$-1 \leq \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \leq 1 \quad \text{(Cauchy Schwarz Inequality)} \quad (10)$$

12. Another example. Because the projection $\hat{y}$ and projection error $\hat{u}$ are orthogonal (perpendicular), the Pythagorean Theorem $c^2 = a^2 + b^2$ implies the following decomposition of total variation of y

$$\text{total sum squares} = \text{explained sum squares} + \text{residual sum squares}$$

and that decomposition implies the R-squared of a regression is between 0 and 1:

$$0 \leq \text{R Squared} \leq 1$$