Linear Algebra Review

Many slides in the first half of this review are from O. Camps (Penn State University)

Why do we need Linear Algebra?

- We will associate coordinates to
  - 3D points in the scene
  - 2D points in the image
- Coordinates will be used to
  - Perform geometrical transformations
  - Associate different coordinate systems (platform, end-effector, camera)
  - Associate 3D with 2D points
- Images are matrices of numbers
  - We will find properties of these numbers

Matrices

$$A_{n\times m} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ a_{31} & a_{32} & \dots & a_{3m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \quad B_{n\times m} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ b_{31} & b_{32} & \dots & b_{3m} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix}$$

Sum:

$$C_{n\times m} = A_{n\times m} + B_{n\times m}$$

$$c_{ij} = a_{ij} + b_{ij}$$

A and B must have the same dimensions

Example:

$$\begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 4 & 6 \end{bmatrix}$$

Matrices

Product:

$$C_{n\times p} = A_{n\times m} B_{m\times p}$$

A and B must have compatible dimensions

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$$

$$A_{n\times n} B_{n\times n} \neq B_{n\times n} A_{n\times n}$$

Examples:

$$\begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 17 & 29 \\ 19 & 11 \end{bmatrix}$$

$$\begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} \cdot \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 18 & 32 \\ 17 & 10 \end{bmatrix}$$

Matrices

Transpose:

$$C_{m\times n} = A^T{}_{n\times m} \qquad (A+B)^T = A^T + B^T$$

$$c_{ij} = a_{ji} \qquad (AB)^T = B^T A^T$$

If $A^T = A$ A is symmetric

Examples:

$$\begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 \\ 2 & 5 \end{bmatrix} \qquad \begin{bmatrix} 6 & 2 \\ 1 & 5 \\ 3 & 8 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 & 3 \\ 2 & 5 & 8 \end{bmatrix}$$

Matrices

Determinant:

A must be square

$$\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Example:

$$\det\begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} = 2 - 15 = -13$$

Matrices

Inverse:

A must be square

$$A_{n\times n} A^{-1}{}_{n\times n} = A^{-1}{}_{n\times n} A_{n\times n} = I$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

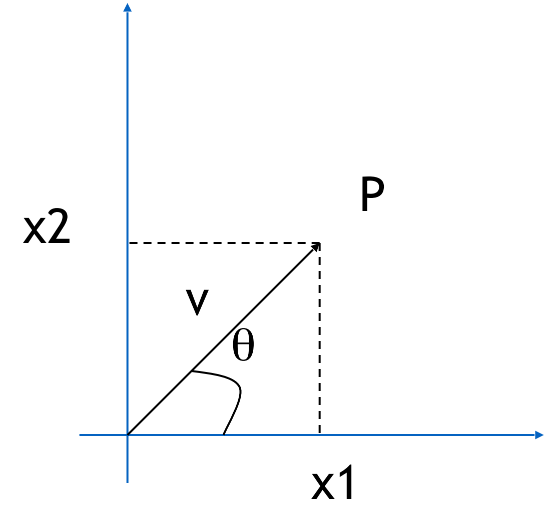
Example:

$$\begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{28} \begin{bmatrix} 5 & -2 \\ -1 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix}^{-1} . \begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} = \frac{1}{28} \begin{bmatrix} 5 & -2 \\ -1 & 6 \end{bmatrix} . \begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} = \frac{1}{28} \begin{bmatrix} 28 & 0 \\ 0 & 28 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2D Vector

$$\mathbf{v} = (x_1, x_2)$$



Magnitude: $$\|\mathbf{v}\| = \sqrt{x_1^{\,2} + x_2^{\,2}}$$

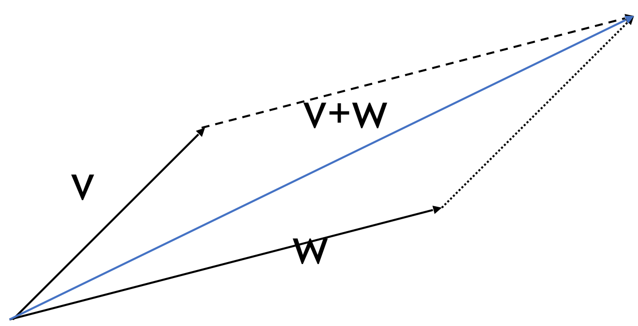
If $\|\mathbf{v}\| = 1$, $\mathbf{v}$ Is a UNIT vector

$$\frac{\mathbf{v}}{\|\mathbf{v}\|} = \left(\frac{x_1}{\|\mathbf{v}\|}, \frac{x_2}{\|\mathbf{v}\|} \right)$$ Is a unit vector

Orientation: $$\theta = \tan^{-1}\left(\frac{x_2}{x_1} \right)$$

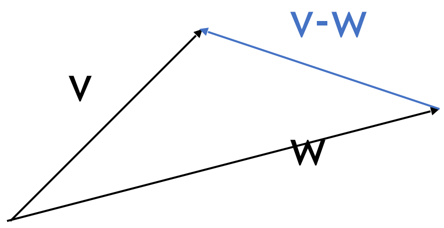
Vector Addition

$$\mathbf{v} + \mathbf{w} = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$$



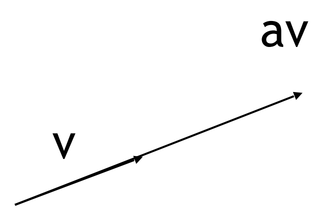
Vector Subtraction

$$\mathbf{v} - \mathbf{w} = (x_1, x_2) - (y_1, y_2) = (x_1 - y_1, x_2 - y_2)$$

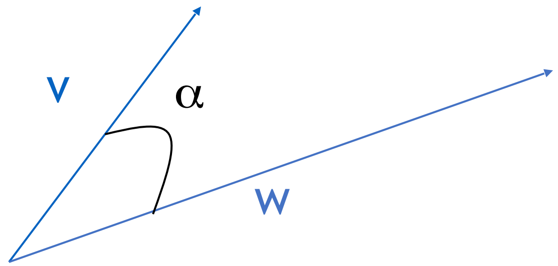


Scalar Product

$$a\mathbf{v} = a(x_1, x_2) = (ax_1, ax_2)$$



Inner (dot) Product



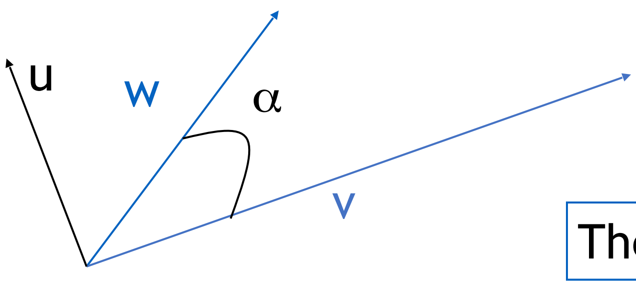
$$v \cdot w = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_1 + x_2 y_2$$

The inner product is a SCALAR!

$$v \cdot w = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = ||v|| \cdot ||w|| cos\, \alpha$$

$$v.w = 0 \Leftrightarrow v \perp w$$

Vector (cross) Product



$$u = v \times w = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}$$

The cross product is a VECTOR!

Magnitude: $\| \mathrm{u} \| = \| v.w \| = \| v \| \| w \| \sin\alpha$

Orientation:

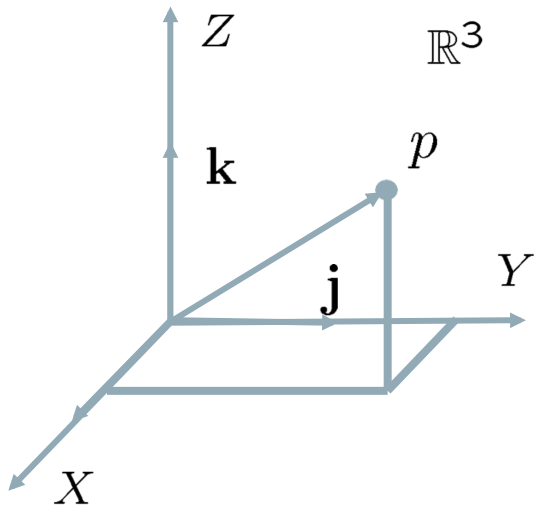
$$u \perp v \Rightarrow u \cdot v = (v \times w) \cdot v = 0$$

$$u \perp w \Rightarrow u \cdot w = (v \times w) \cdot w = 0$$

Orthonormal Basis in 3D

Standard base vectors:

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$



Coordinates of a point p in space:

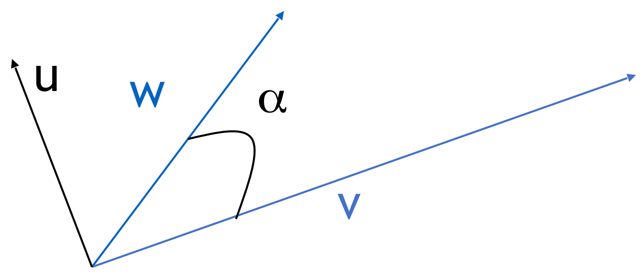
$$\boldsymbol{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \in \mathbb{R}^3 \qquad \boldsymbol{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = X.\mathbf{i} + Y.\mathbf{j} + Z.\mathbf{k}$$

Vector Product Computation

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{u} = \mathbf{v} \times \mathbf{w} = (x_1, x_2, x_3) \times (y_1, y_2, y_3)$$

$$\mathbf{u} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$



$$= (x_2 y_3 - x_3 y_2)\mathbf{i} + (x_3 y_1 - x_1 y_3)\mathbf{j} + (x_1 y_2 - x_2 y_1)\mathbf{k}$$

Least Squares Optimization

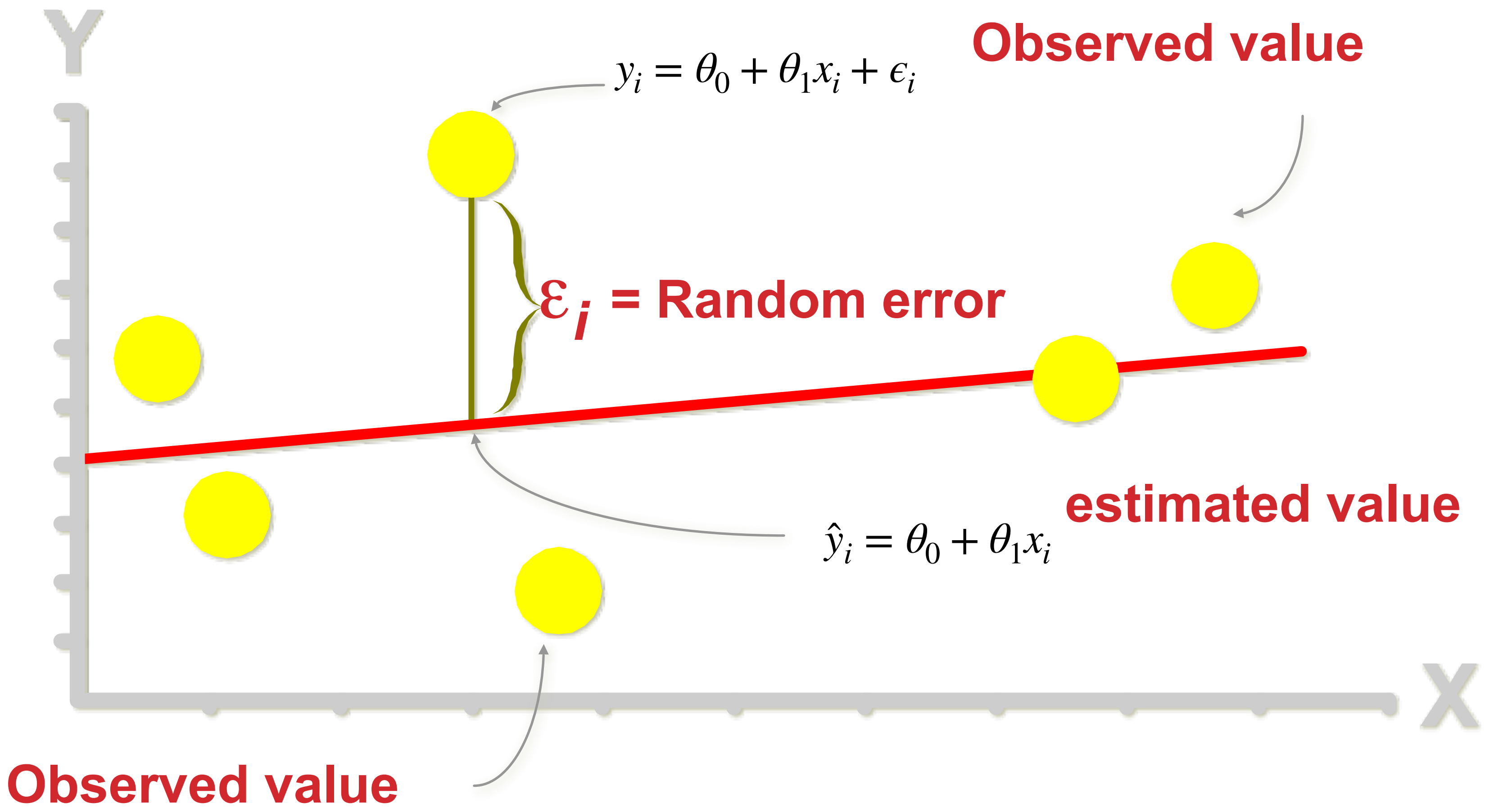
Example

- Single Variable Linear Regression

estimate $\hat{y}_i = \theta_0 + \theta_1 x_i$

x Area(sq. ft.)	y Price (in 1000$)
1600	220
1400	180
2100	350
…	…
….	….
2400	500

LINEAR REGRESSION



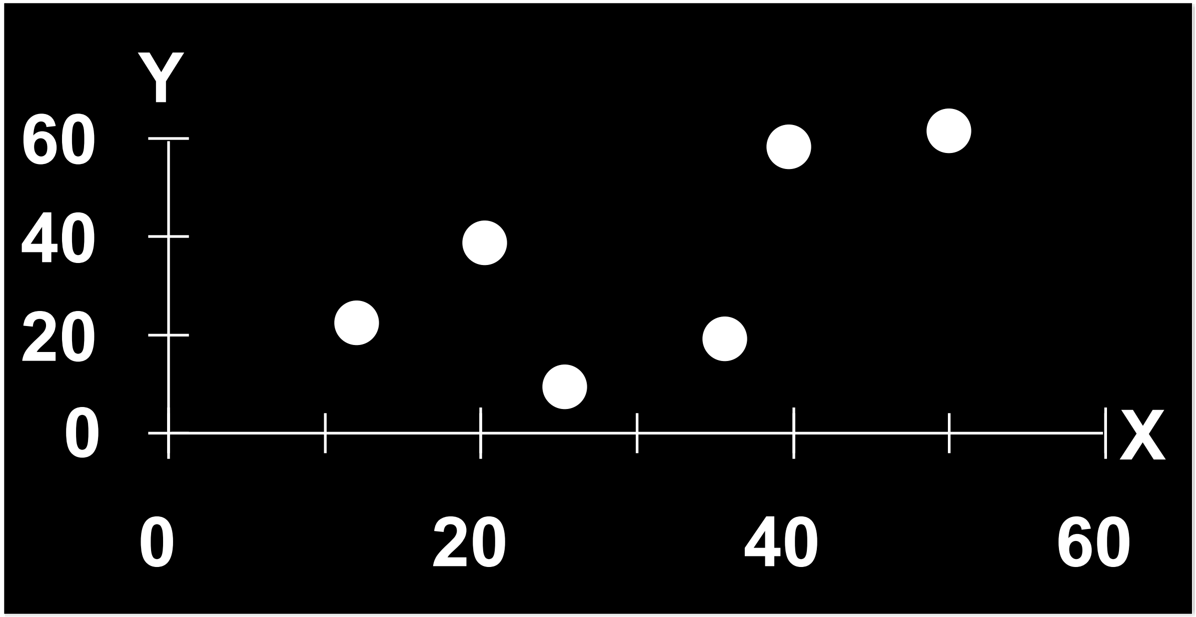
[WF]

?

ESTIMATING PARAMETERS: LEAST SQUARES METHOD

SCATTER PLOT

Plot all (X_i, Y_i) pairs, and plot your learned model



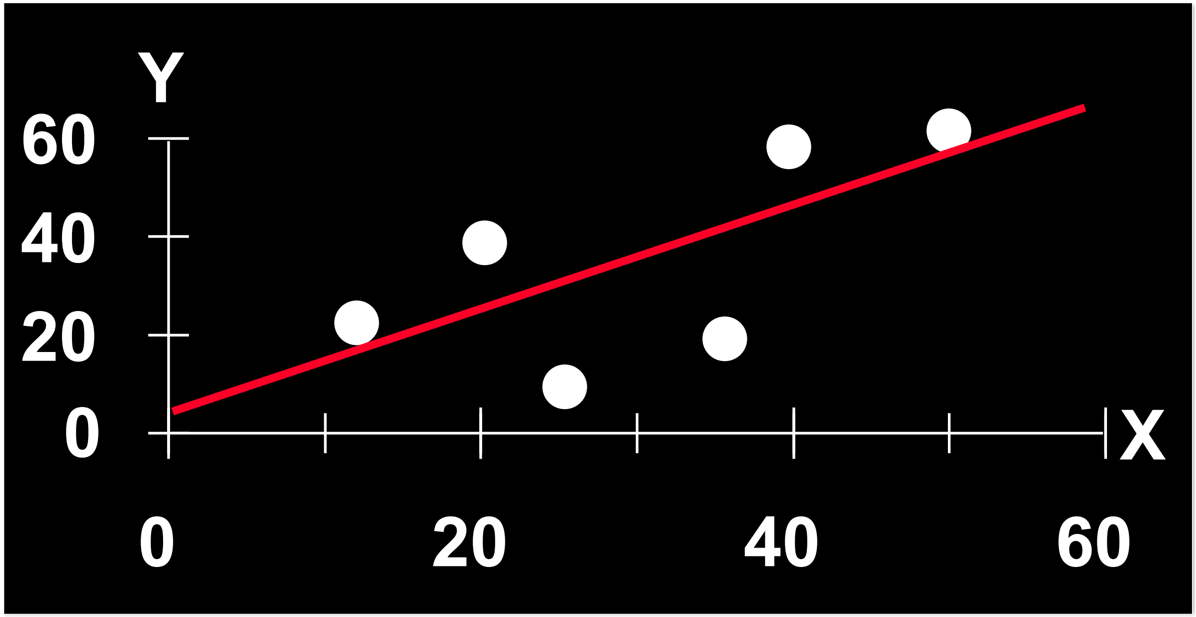
[WF]

QUESTION

How would you draw a line through the points?

How do you determine which line "fits the best" …?

??????????

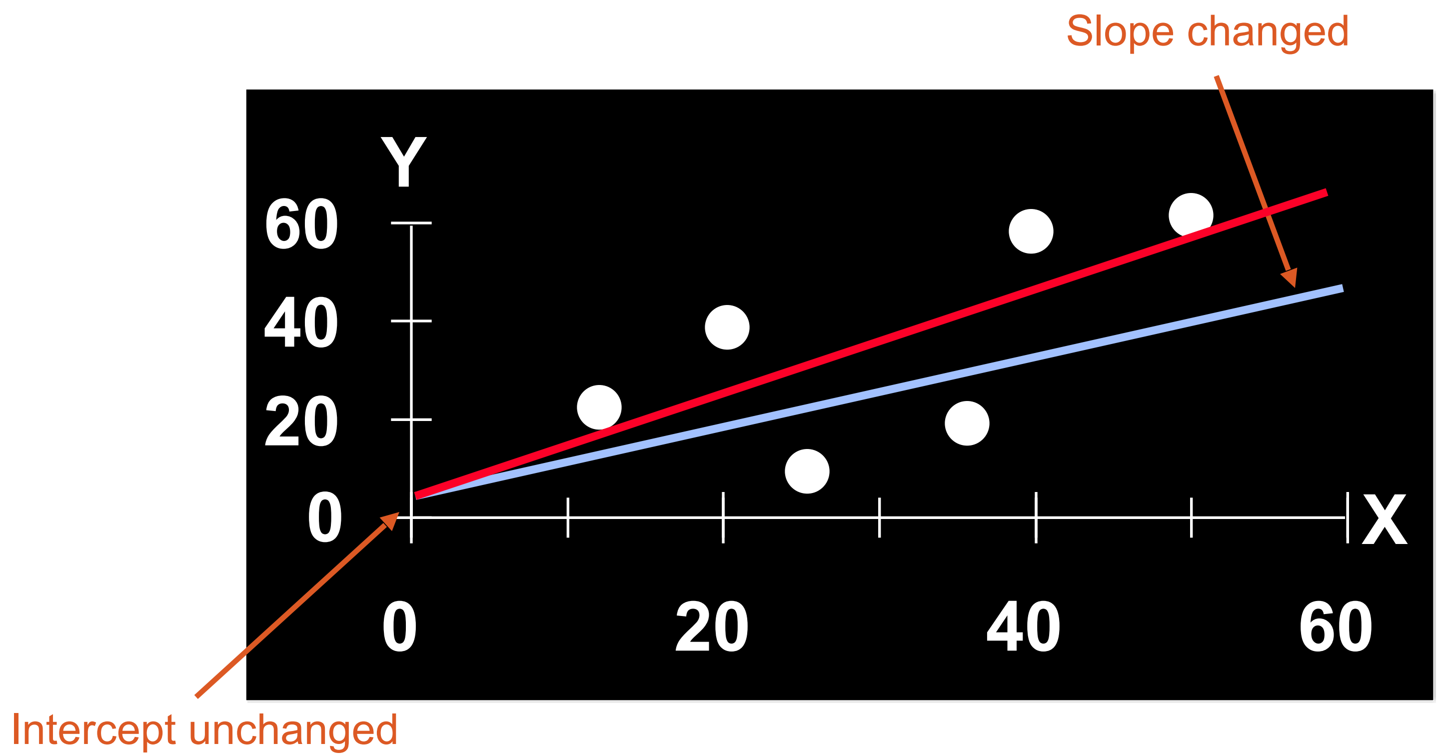


[WF]

QUESTION

How would you draw a line through the points?

How do you determine which line "fits the best" ?????????



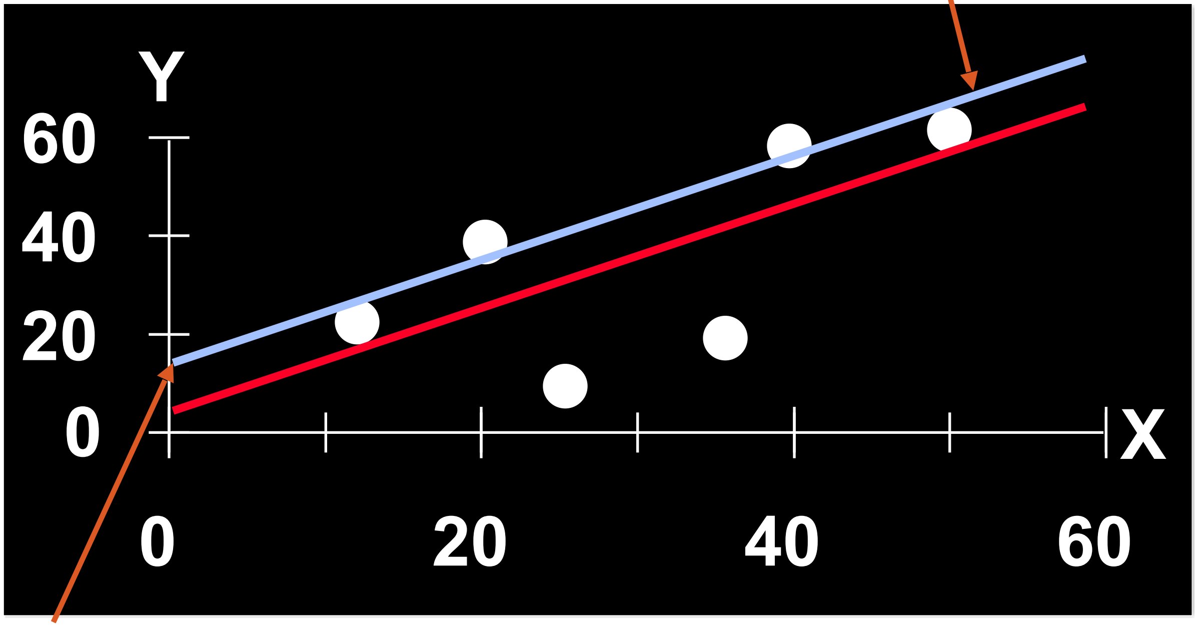
[WF]

QUESTION

How would you draw a line through the points?

How do you determine which line "fits the best" ?????????

Slope unchanged



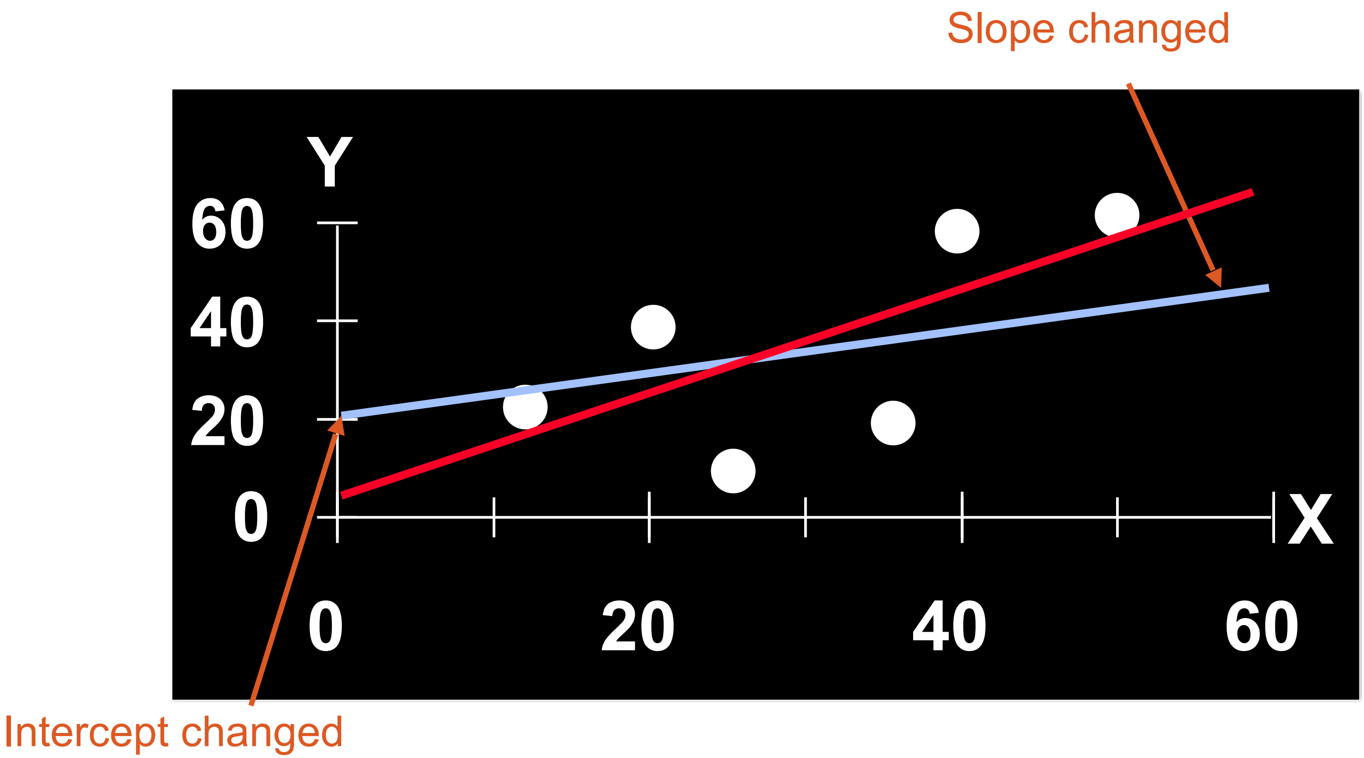
Intercept changed

[WF]

QUESTION

How would you draw a line through the points?

How do you determine which line "fits the best" ?????????



[WF]

LEAST SQUARES

Best fit: difference between the true (observed) Y-values and the estimated Y-values is minimized:

- Positive errors offset negative errors …
- … square the error!

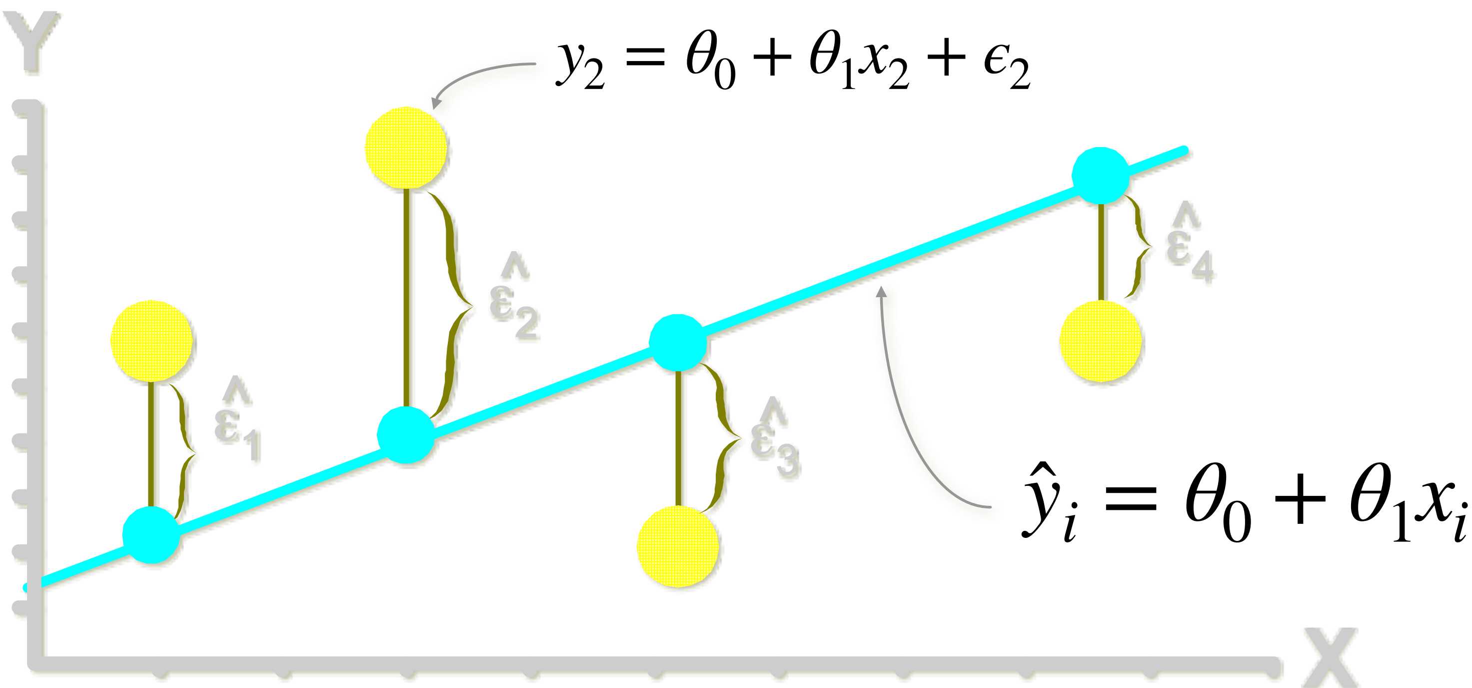
$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \epsilon_i^2$$

Least squares minimizes the sum of the squared errors

[WF]

LEAST SQUARES, GRAPHICALLY

LS Minimizes $\sum_{i=1}^{n} \epsilon_i^2 = \epsilon_1^2 + \epsilon_2^2 + \ldots + \epsilon_n^2$



[WF]

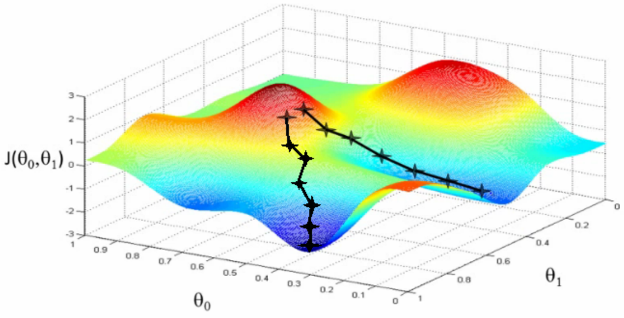
Example

- Single Variable Linear Regression

estimate $\hat{y}_i = \theta_0 + \theta_1 x_i$

x Area(sq. ft.)	y Price (in 1000$)
1600	220
1400	180
2100	350
…	…
….	….
2400	500

Multivariate Regression

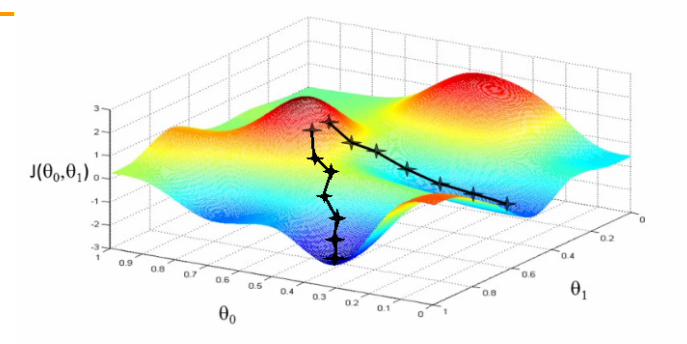


- Multi Linear Regression

$$\hat{y}_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{in}$$

y Price (in 1000$)	x_1 Area(sq. ft.)	x_2 # Bathrooms	x_3 # Bedrooms
220	1600	2.5	3
180	1400	1.5	3
350	2100	3.5	4
…	…	…	…
….	….	…	…
500	2400	4	5

Multivariate Regression



- Multi Linear Regression

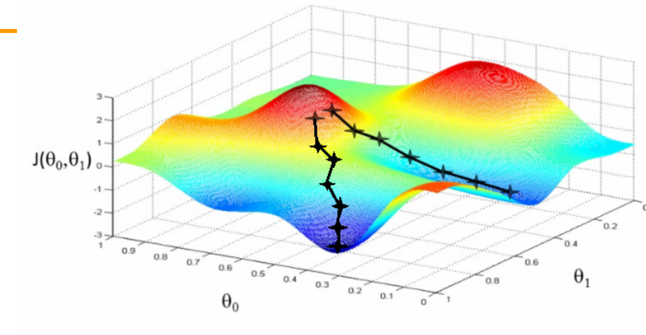
$$\hat{y}_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{in}$$

	Price (in 1000$)	Area(sq. ft.)	# Bathrooms	# Bedrooms
	220	1600	2.5	3
y_i	180	1400	1.5	3
	350	2100	3.5	4
	…	…	…	…
	….	….	…	…
	500	2400	4	5

x_i

1400	x_{i1}
1.5	x_{i2}
3	x_{i3}

Multivariate Regression



- Multi Linear Regression

$$y_i = \theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{in}$$

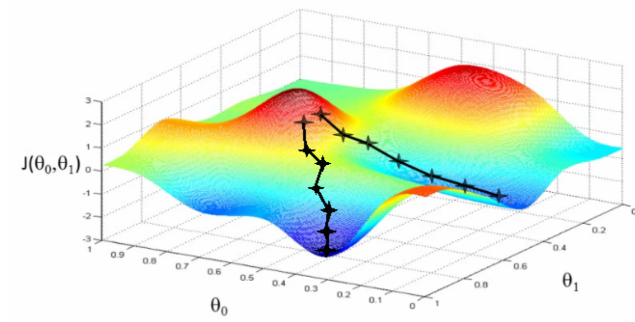
y Price (in 1000$)	x_0	x_1 Area(sq. ft.)	x_2 # Bathrooms	x_3 # Bedrooms
220	1	1600	2.5	3
y_i 180	1	1400	1.5	3
350	1	2100	3.5	4
…	…	…	…	…
….	….	….	…	…
500	1	2400	4	5

x_i

1	x_{i0}
1400	x_{i1}
1.5	x_{i2}
3	x_{i3}

Multivariate Regression Model

- Model:

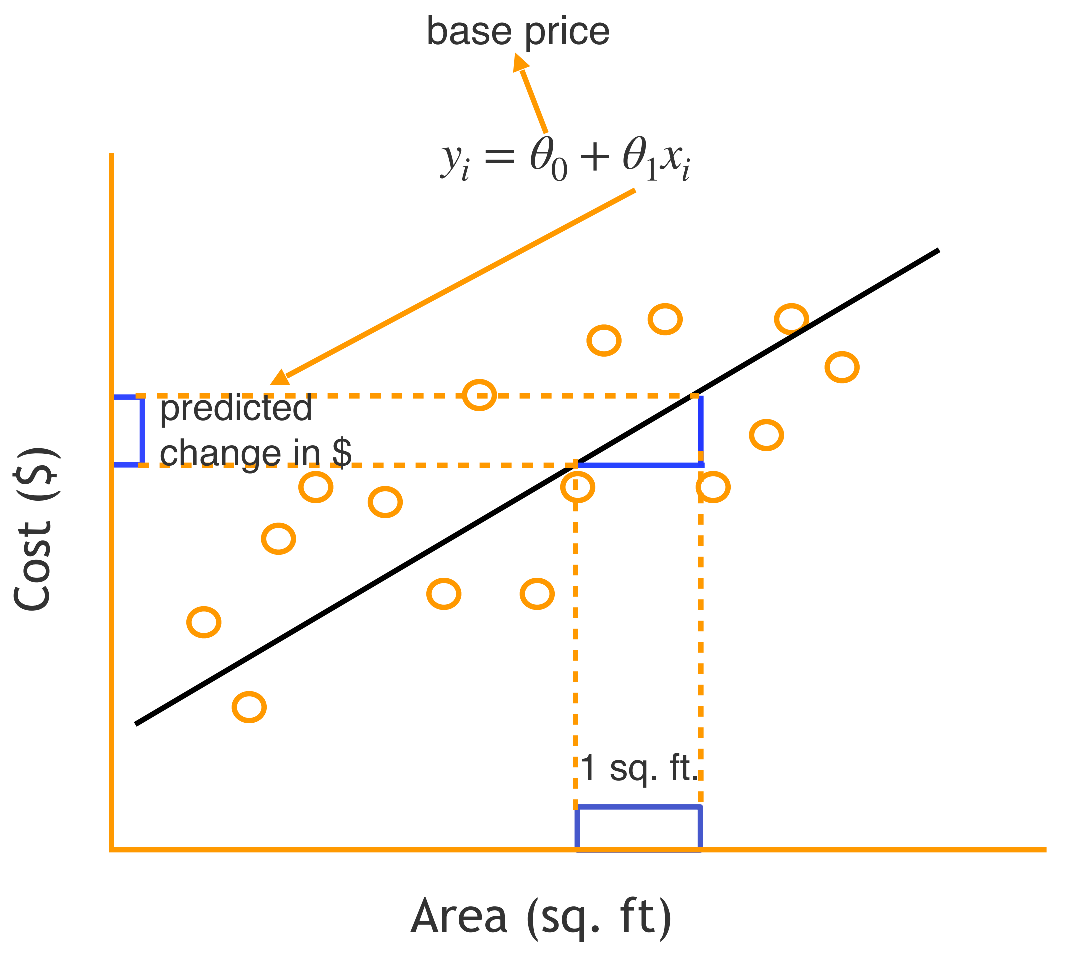


$$\hat{y}_i = \theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{in}$$

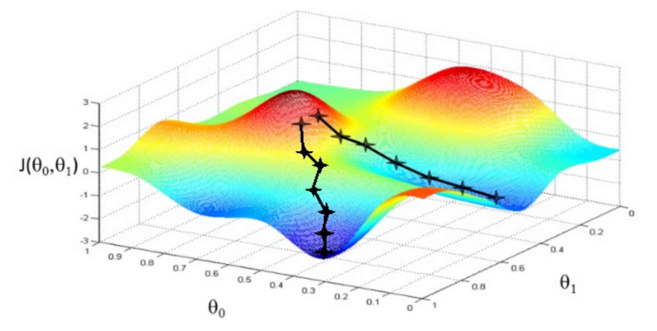
$$\hat{y}_i = \sum_{j=0}^{n} \theta_{ij} x_{ij}$$

feature 1 = x_0 …. (constant, 1)
feature 2 = x_1 …. (area, sq. ft.)
feature 3 = x_2 …. (# of bedrooms)
feature 4 = x_3 …. (# of bathrooms)
….
….
feature n = x_n

Single Variable Linear Regression

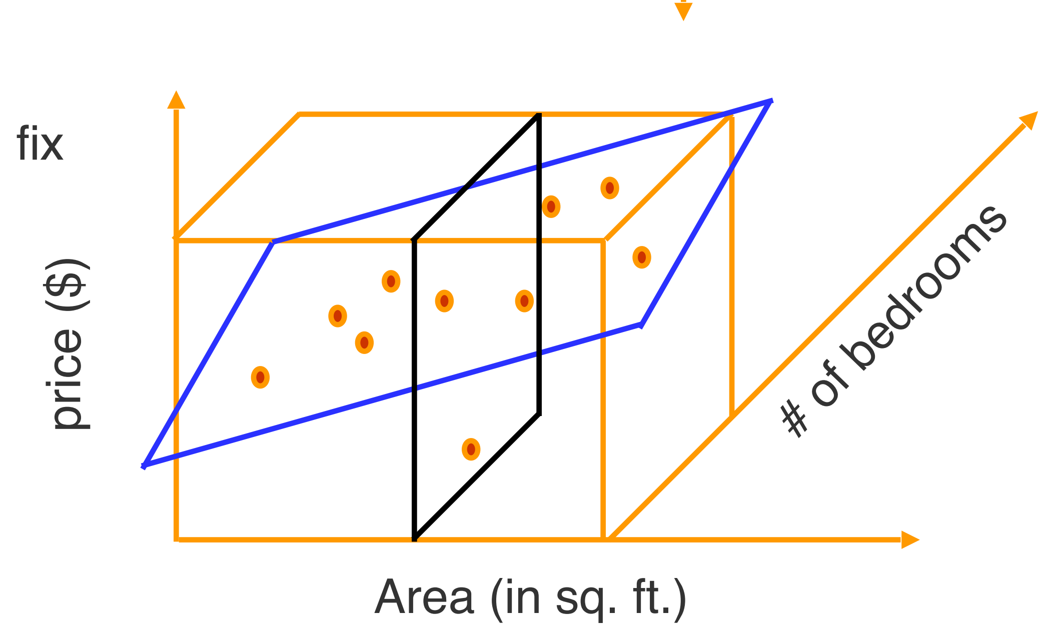


Interpreting Coefficients



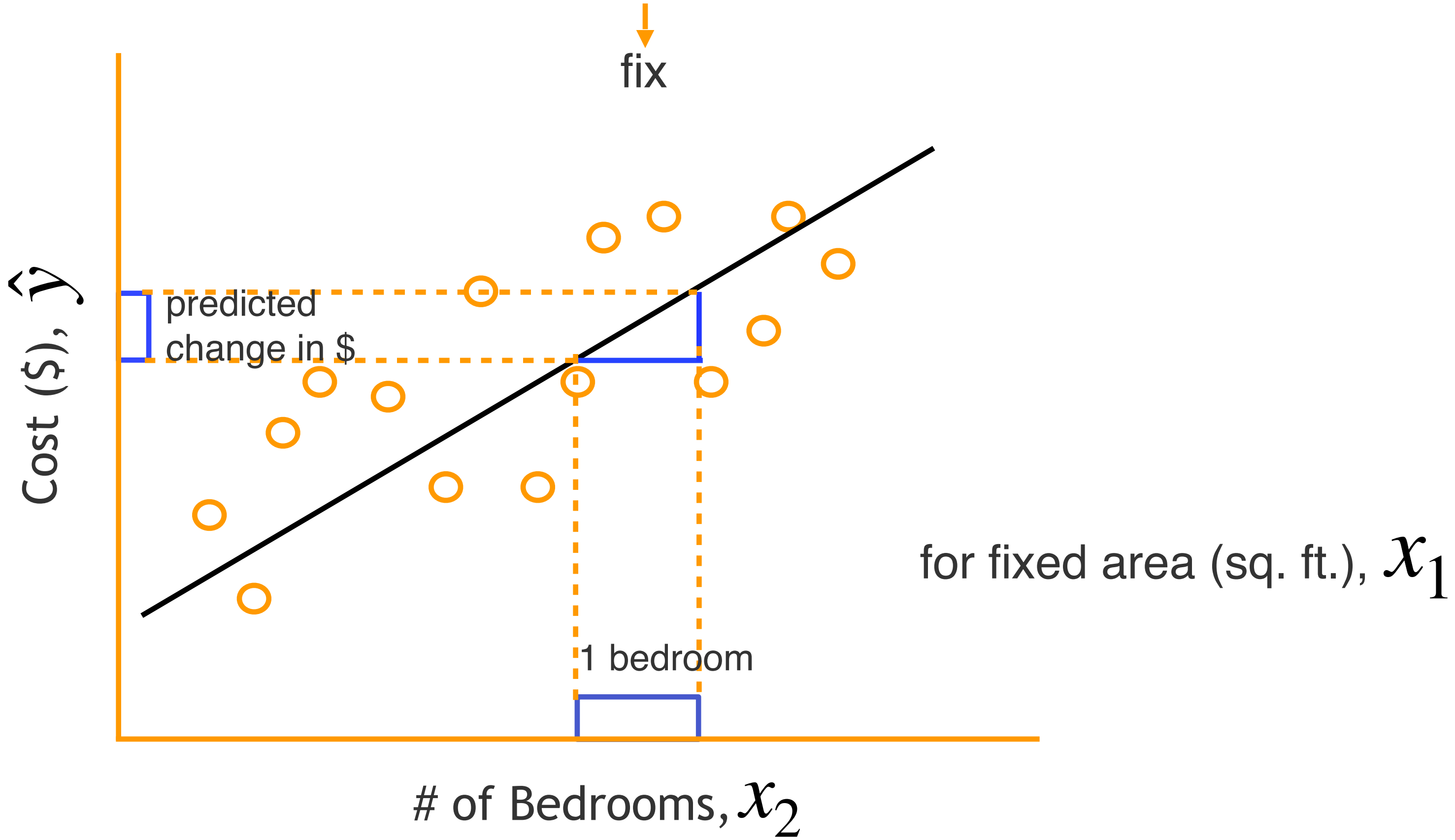
- Two Linear Features

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2$$



Single Variable Linear Regression

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2$$

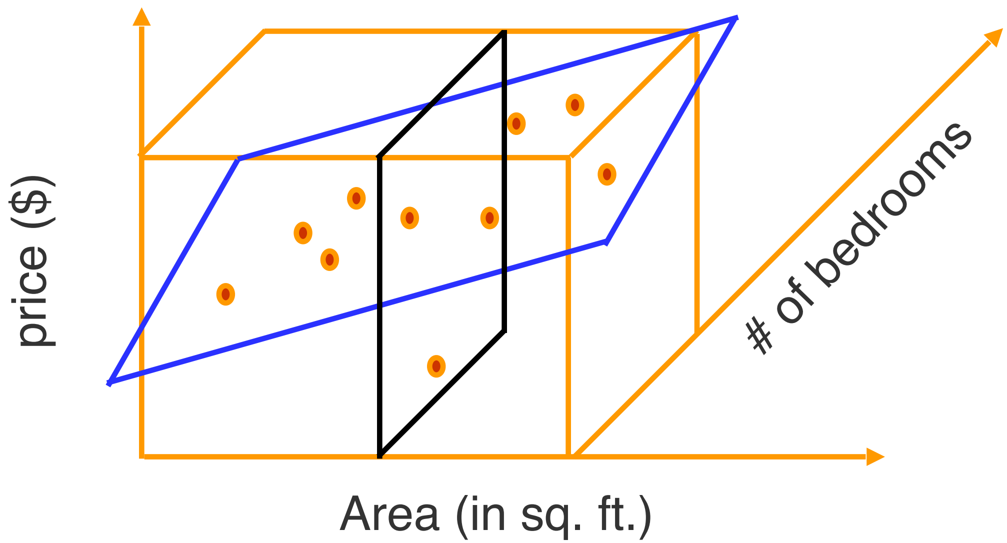


Interpreting Coefficients

- Multiple Features

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \ldots + \hat{\theta}_j x_j + \ldots + \hat{\theta}_m x_m$$

fix fix fix fix

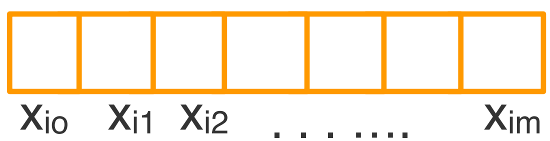


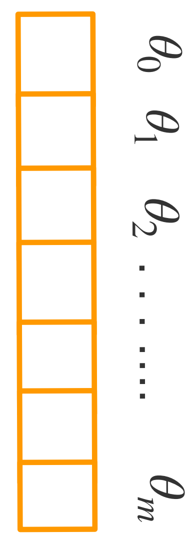
One Observation Model

- Matrix Notation
  For observation i

$$\hat{y}_i = \sum_{j=0}^{m} \theta_{ij} x_{ij}$$

$y_i =$





$$y_i = X_i^T \theta$$

All Observation Model

- Matrix Notation
  For all observations

$$
\begin{bmatrix}
x_{10} & x_{11} & x_{12} & .. & .. & x_{1m} \\
x_{20} & x_{21} & x_{22} & .. & .. & x_{2m} \\
x_{30} & x_{31} & x_{32} & .. & .. & x_{3m} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
x_{n0} & x_{n1} & x_{n2} & .. & .. & x_{nm}
\end{bmatrix}
\begin{bmatrix}
\theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ \theta_m
\end{bmatrix}
=
\begin{bmatrix}
y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n
\end{bmatrix}
$$

$$\hat{Y} = X\theta$$

LEAST SQUARES OPTIMIZATION

Each row is a feature vector paired with a label for a single input

Rewrite inputs:

n labeled inputs

$$X = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \cdots \\ (x^{(n)})^T \end{bmatrix} \in \mathbb{R}^{n \times m}, y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \cdots \\ y^{(n)} \end{bmatrix} \in \mathbb{R}^n$$

m features

Rewrite optimization problem:

$$\text{minimize}_\theta \ \frac{1}{2}\|X\theta - y\|_2^2$$

*Recall $||z||_2^2 = z^T z = \sum z_i^2$

LEAST SQUARES OPTIMIZATION

Rewrite inputs:

Each row is a feature vector paired with a label for a single input

n labeled inputs

$$X = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \cdots \\ (x^{(n)})^T \end{bmatrix} \in \mathbb{R}^{n \times m}, y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \cdots \\ y^{(n)} \end{bmatrix} \in \mathbb{R}^n$$

m features

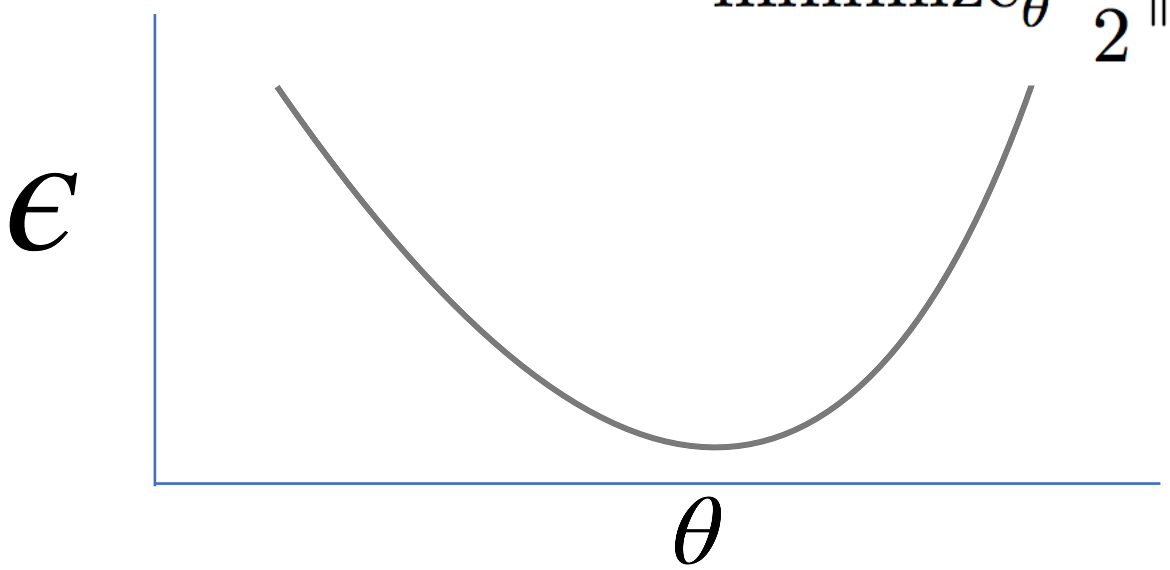
Rewrite optimization problem: $\text{minimize}_\theta \ \frac{1}{2}\|X\theta - y\|_2^2$

$$\implies \text{minimize} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \epsilon_i^2$$

*Recall $||z||_2^2 = z^T z = \sum z_i^2$

ERROR FUNCTION

$$\text{minimize}_\theta \; \frac{1}{2}\|X\theta - y\|_2^2$$



$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\epsilon_i^2$$

GRADIENTS

Minimizing a multivariate function involves finding a point where the gradient is zero:

$$\nabla_\theta f(\theta) = 0 \text{ (the vector of zeros)}$$

Points where the gradient is zero are local minima

- If the function is convex, also a global minimum

Let's solve the least squares problem!

We'll use the multivariate generalizations of some concepts from MATH141/142 …

- Chain rule: $\nabla_\theta f(X\theta) = X^T \nabla_{X\theta} f(X\theta)$
- Gradient of squared ℓ^2 norm: $\nabla_\theta \|\theta - z\|_2^2 = 2(\theta - z)$

LEAST SQUARES

Recall the least squares optimization problem:

$$\text{minimize}_\theta \ \frac{1}{2}\|X\theta - y\|_2^2$$

What is the gradient of the optimization objective ????????

$$\nabla_\theta \frac{1}{2}\|X\theta - y\|_2^2 =$$

Chain rule:

$$\nabla_\theta f(X\theta) = X^T \nabla_{X\theta} f(X\theta)$$

$$X^T \nabla_{X\theta} \frac{1}{2}\|X\theta - y\|_2^2 =$$

Gradient of norm:

$$\nabla_\theta \|\theta - z\|_2^2 = 2(\theta - z)$$

$$\nabla_\theta \frac{1}{2}\|X\theta - y\|_2^2 = X^T(X\theta - y)$$

LEAST SQUARES

Recall: points where the gradient equals zero are minima.

$$\nabla_\theta \frac{1}{2}\|X\theta - y\|_2^2 = X^T(X\theta - y)$$

So where do we go from here?????????

Solve for model parameters θ

$$X^T(X\theta - y) = 0$$

$$X^TX\theta - X^Ty = 0 \Rightarrow X^TX\theta = X^Ty$$

$$(X^TX)^{-1}X^TX\theta = (X^TX)^{-1}X^Ty$$

$$\theta = (X^TX)^{-1}X^Ty$$