Least squares regression

Data:

	dependent x	independent y
(1)	$x^{(1)}$	$y^{(1)}$
(2)	$x^{(2)}$	$y^{(2)}$
(3)	$x^{(3)}$	$y^{(3)}$

Model: $y^{(i)} = p\, x^{(i)} + \epsilon^{(i)}$

↳ p: scale factor

↳ ϵ: error

$$\vec{y} = p\,\vec{x} + \vec{\epsilon}$$

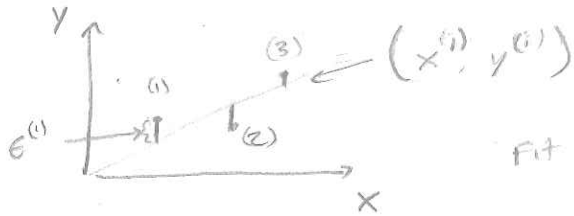
Problem:

Which p will minimize

$$\sum_n \left(\epsilon^{(n)}\right)^2 = \|\vec{\epsilon}\|^2$$

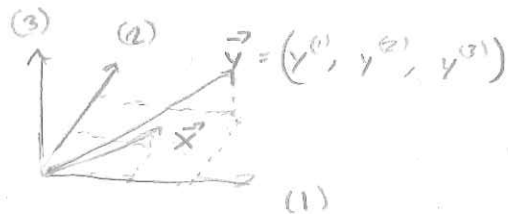
squared errors

Picture 1: Each datapoint is a vector

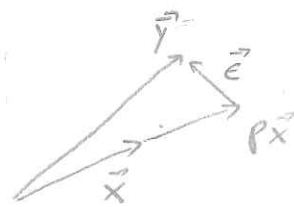


Fit a line

Picture 2: $\vec{x}$ is one vector, $\vec{y}$ is another



What are p and $\vec{\epsilon}$ in picture 2?



$$\vec{y} = p\vec{x} + \vec{\epsilon}$$

Solve for p:

1) As the length of $\vec{y}$ projected onto $\hat{x}$:

$$\|p\,\vec{x}\| = \vec{y}\cdot\hat{x}$$

$$p\,\vec{x} = (\vec{y}\cdot\hat{x})\,\hat{x}$$

↳ $\hat{x} = \dfrac{\vec{x}}{\|\vec{x}\|}$

$$p\vec{x} = \left(\vec{y}\cdot\frac{\vec{x}}{\|x\|}\right)\frac{\vec{x}}{\|x\|}$$

$$p = \frac{\vec{y}\cdot\vec{x}}{\|x\|^2} = \frac{\vec{y}\cdot\vec{x}}{\vec{x}\cdot\vec{x}} = \frac{y^T x}{x^T x}$$

2) As the value that makes $\vec{\epsilon}$ orthogonal to $\vec{x}$:

$$\vec{x}\cdot\vec{\epsilon} = 0$$

$$\vec{x}\cdot(\vec{y} - p\vec{x}) = 0$$

$$\vec{x}\cdot\vec{y} - p(\vec{x}\cdot\vec{x}) = 0$$

$$p = \frac{\vec{y}\cdot\vec{x}}{\vec{x}\cdot\vec{x}}$$

Multiple regression

two regressors

Data:

$$\begin{array}{c|c|cc} & y & x_1 & x_2 \\ (1) & \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{bmatrix} & \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ x_1^{(3)} & x_2^{(3)} \end{bmatrix} \\ (2) & & \\ (3) & & \end{array}$$

Model:

$$\vec{y} = X\vec{p} + \vec{\epsilon}$$

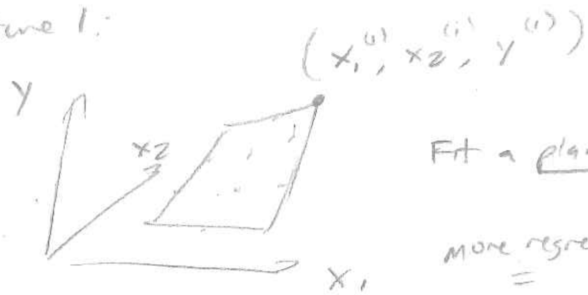
$\vec{\epsilon}$ ↳ one per data point

$\vec{p}$ ↳ one per regressor

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{bmatrix} = \underbrace{\begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ x_1^{(3)} & x_2^{(3)} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}} + \begin{bmatrix} \epsilon^{(1)} \\ \epsilon^{(2)} \\ \epsilon^{(3)} \end{bmatrix}$$

$$X\vec{p} = \begin{bmatrix} p_1 x_1^{(1)} + p_2 x_2^{(1)} \\ p_1 x_1^{(2)} + p_2 x_2^{(2)} \\ p_1 x_1^{(3)} + p_2 x_2^{(3)} \end{bmatrix}$$

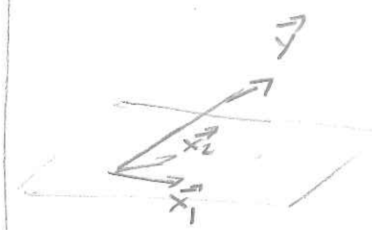
Picture 1:



Fit a plane

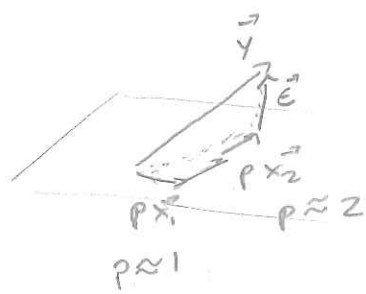
more regressors = more axes

Picture 2:



More regressors = More vectors

What are p and $\vec{\epsilon}$?



Solve for $\vec{p}$ as the value that makes $\vec{\epsilon}$ perpendicular to every $\vec{x_i}$:

$$x_i^T(\vec{y} - p_i\vec{x_i}) = 0$$

$$X^T(\vec{y} - X\vec{p}) = \vec{0}$$

$$X^T\vec{y} - X^TX\vec{p} = \vec{0}$$

$$X^T\vec{y} = X^TX\vec{p}$$

$$(X^TX)^{-1}X^T\vec{y} = \vec{p}$$

What happens if ....

1) All regressors are orthogonal?
2) Not totally orthogonal regressors?
3) Regressors are linearly dependent?

First, what is $X^T X$?

$$X^T X = \begin{bmatrix} x_1 & x_1 & x_1 \\ x_2 & x_2 & x_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_1 & x_2 \\ x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 \\ x_2 \cdot x_1 & x_2 \cdot x_2 \end{bmatrix}$$

---

1) Orthogonal regressors, $x_1 \cdot x_2 = 0$

$$X^T X = \begin{bmatrix} \|x_1\|^2 & 0 \\ 0 & \|x_2\|^2 \end{bmatrix} \qquad (X^T X)^{-1} = \begin{bmatrix} \frac{1}{\|x_1\|^2} & 0 \\ 0 & \frac{1}{\|x_2\|^2} \end{bmatrix}$$

$$\vec{p} = (X^T X)^{-1} X^T \vec{y}$$

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\|x_1\|^2} & 0 \\ 0 & \frac{1}{\|x_2\|^2} \end{bmatrix} \underbrace{\begin{bmatrix} x_1 & x_1 & x_1 \\ x_2 & x_2 & x_2 \end{bmatrix} \begin{bmatrix} y \\ y \\ y \end{bmatrix}}_{\begin{bmatrix} \vec{x}_1 \cdot y \\ \vec{x}_2 \cdot y \end{bmatrix}} \longrightarrow p_1 = \frac{\vec{x}_1 \cdot \vec{y}}{\|x_1\|^2}$$

same as single regression!
x_1 weight does not depend on x_2 if $x_1 \perp x_2$

---

2) Not totally orthogonal regressors

$$(X^T X)^{-1} = \frac{1}{\text{stuff}} \begin{bmatrix} x_2 \cdot x_2 & -x_2 \cdot x_1 \\ -x_1 \cdot x_2 & x_1 \cdot x_1 \end{bmatrix}$$

$$\vec{p} = \underbrace{\quad '' \quad}_{} \begin{bmatrix} \vec{x}_1 \cdot \vec{y} \\ \vec{x}_2 \cdot \vec{y} \end{bmatrix}$$

$$p_1 = \frac{x_2 \cdot x_2}{\text{stuff}} \vec{x}_1 \cdot \vec{y} - \frac{x_2 \cdot x_1}{\text{stuff}} \vec{x}_2 \cdot \vec{y}$$

---

3) Linearly dependent regressors $\quad x_2 = c\, x_1$

$$X^T X = \begin{bmatrix} x_1 \cdot x_1 & c\,(x_1 \cdot x_1) \\ c(x_1 \cdot x_1) & c^2 (x_1 \cdot x_1) \end{bmatrix} \longrightarrow$$

This regression is unconstrained,
The $X^T X$ matrix is not invertible.
Its nullspace is nonzero.

Related by a factor of c!