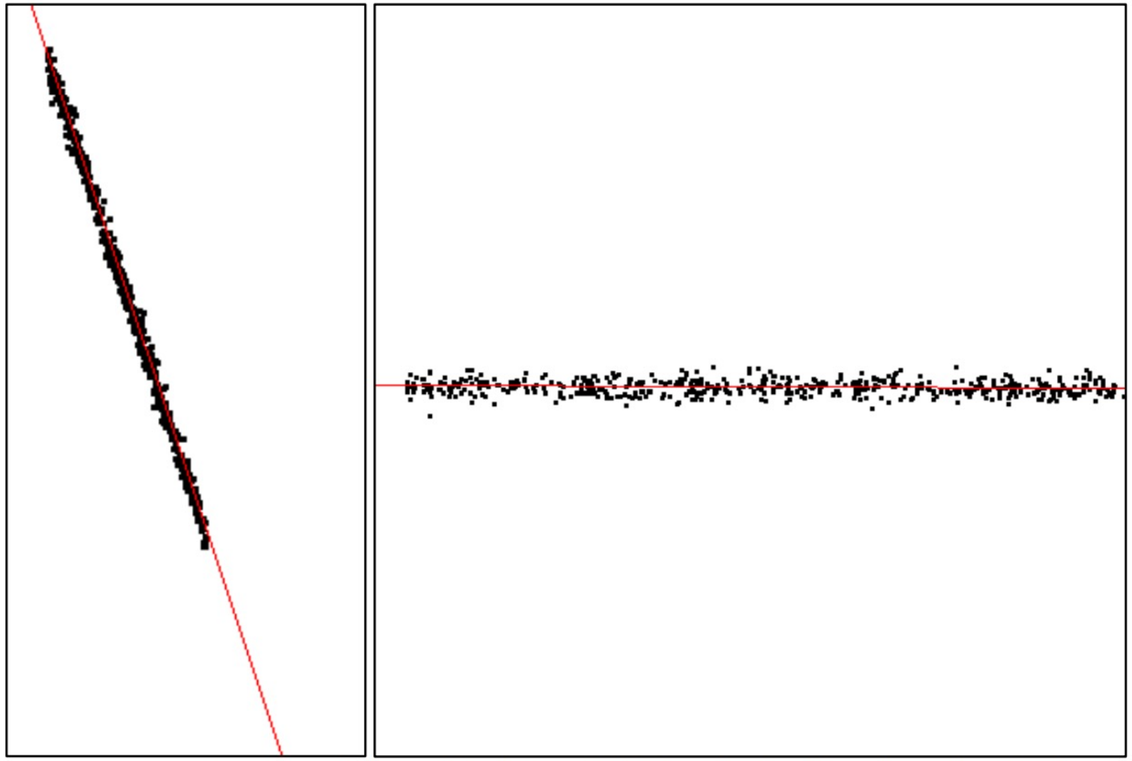
SRF L2

Metoda celor mai mici pătrate

Metoda celor mai mici pătrate

Obiectiv: potrivirea unei linii la o mulţime de puncte 2D
Metoda: regresia liniara cu metoda celor mai mici pătrate (Least Square Fitting)

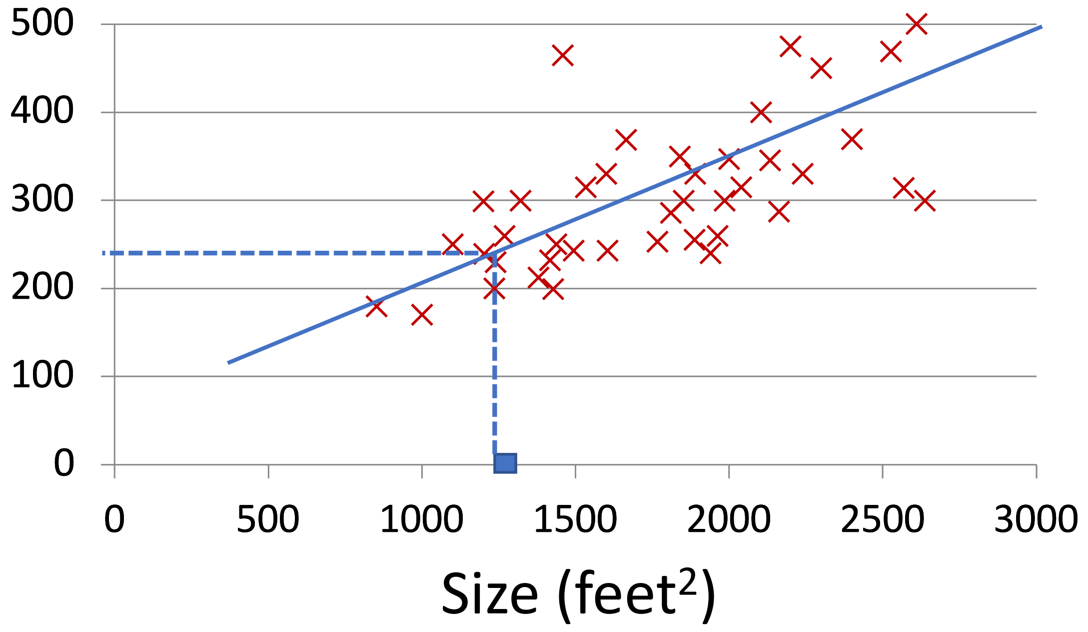
Input:
o mulţime de puncte bidimensionale de forma (xi , yi) unde i = $\{1,2, \dots , n\}$.
Output: ecuatia dreptei care se potriveste cel mai bine la aceste puncte



Least Mean Squares - Application

Housing Prices

Price
(in 1000s
of dollars)



What’s the price of a house having 1250 feet^2 ?

1. Fit a line f(x) to the points
2. Make a prediction for x = 1250 => f(1250) = 250

Least Mean Squares

<u>Supervised Learning</u>

Given the "right answer" for each example in the data (training set)

<u>Regression Problem</u>

Predict real-valued output for new data

Notation:

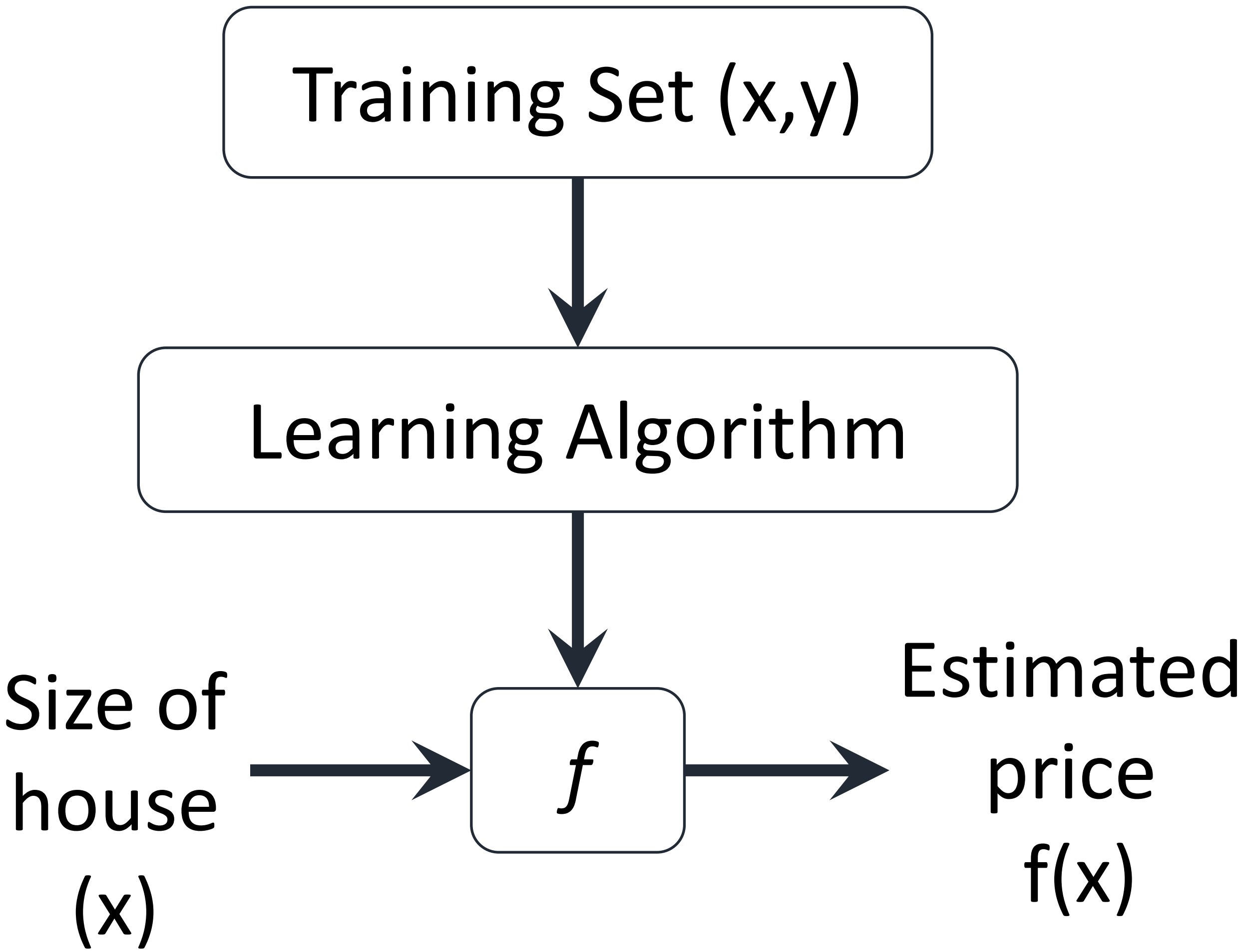
m = Number of training examples
x's = "input" variable / features
y's = "output" variable / "target" variable

(x, y) – one training example

Training set of housing prices

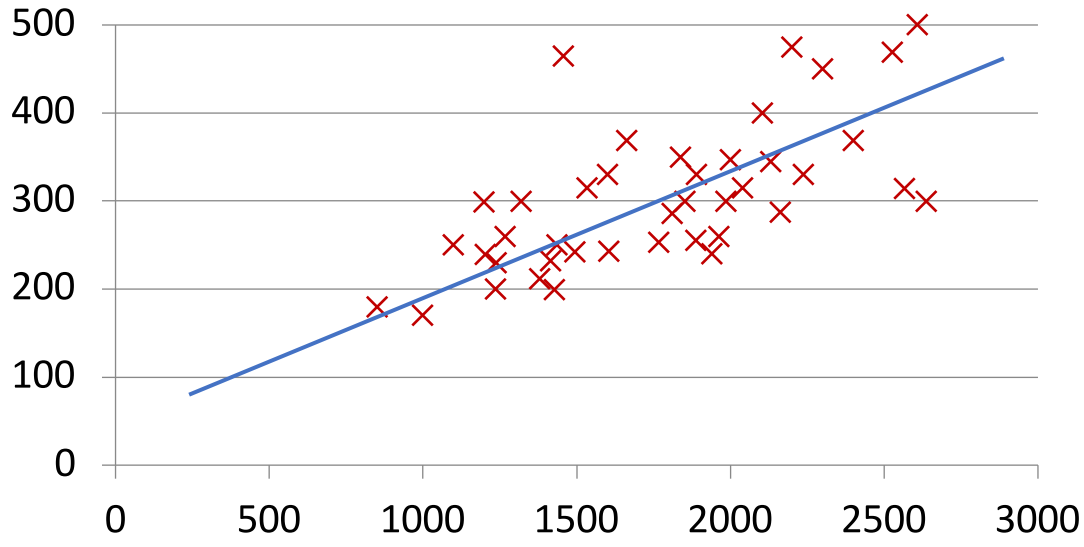
Size in feet^2 (x)	**Price ($) in 1000's** (y)
2104	460
1416	232
1534	315
852	178
…	…

Least Mean Squares



How do we represent *f* ?

Eq of a line: $f(x) = \theta_0 + \theta_1 x$

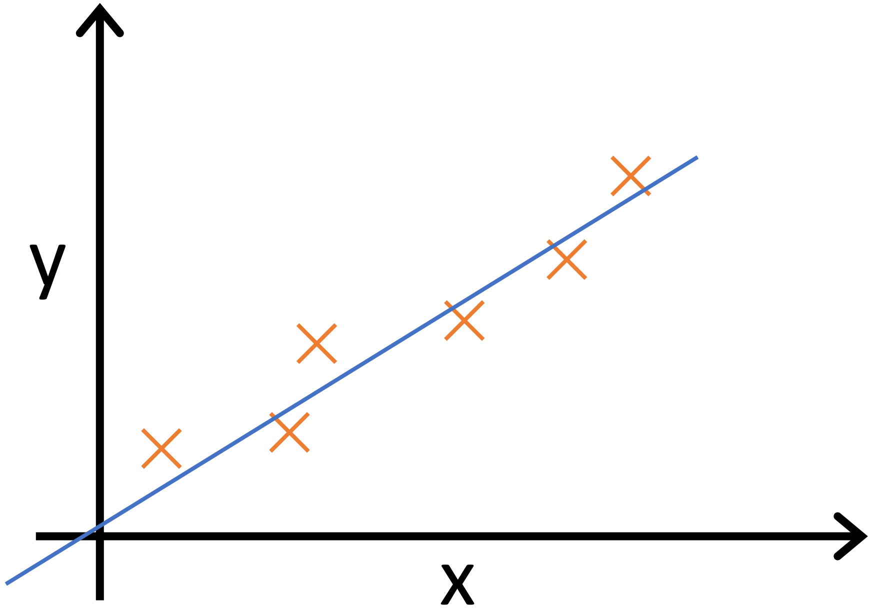


Model 1

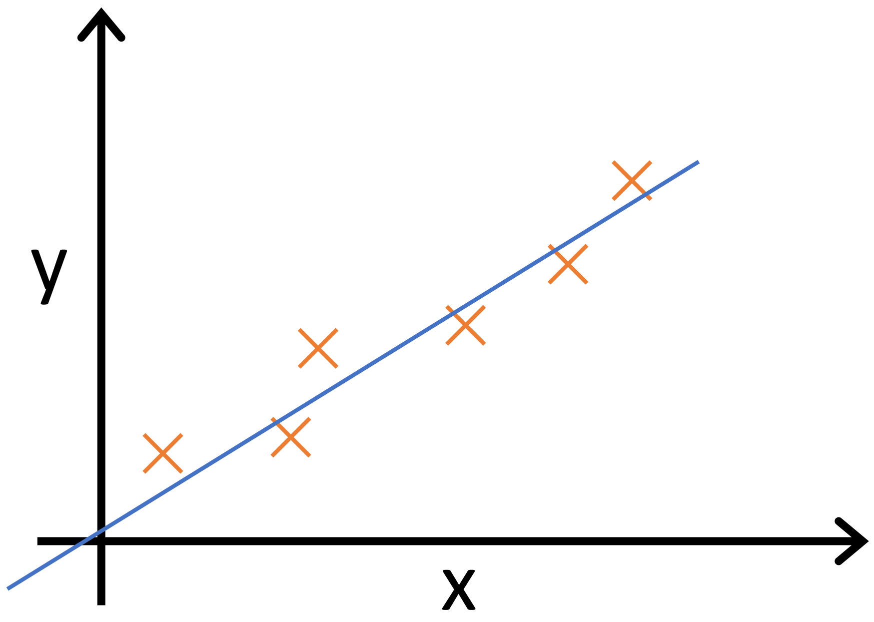
Model 1 – equation of a line: $f(x) = \theta_0 + \theta_1 x$

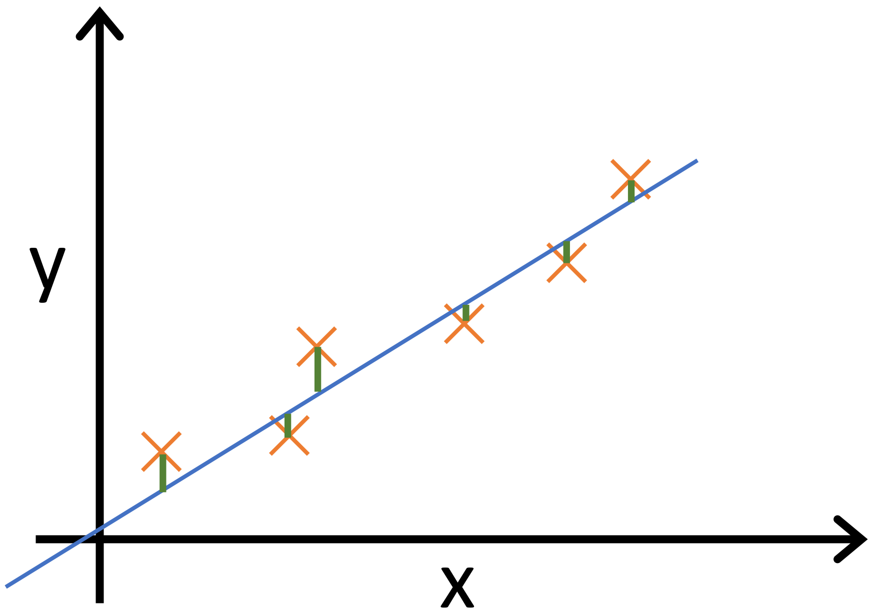
- Slope-intercept form (slope = θ_1 , intercept = θ_0)
- Given the training set with points (xi , yi) , how to find the θ_i parameters of the line f(x) that best fits the points?

Model 1



Idea: Choose θ_0, θ_1 so that
f(x) is close to y for our
training examples (x, y)





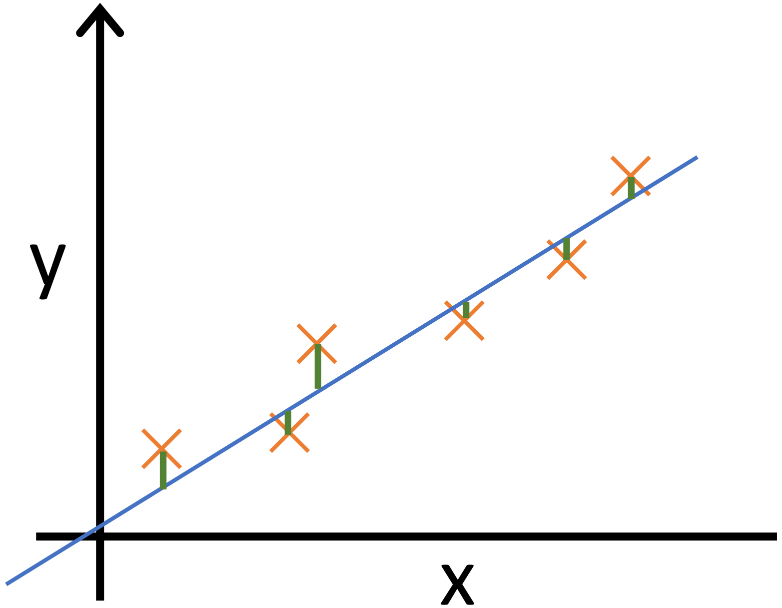
The green line shows distance from the point to the line on y axis

Idea: Choose θ_0, θ_1 so that f(x) is close to y for our training examples (x, y)

$$f(x_i) - y$$

is the distance from point $\mathbf{x_i}$ to the line on the **y** axis

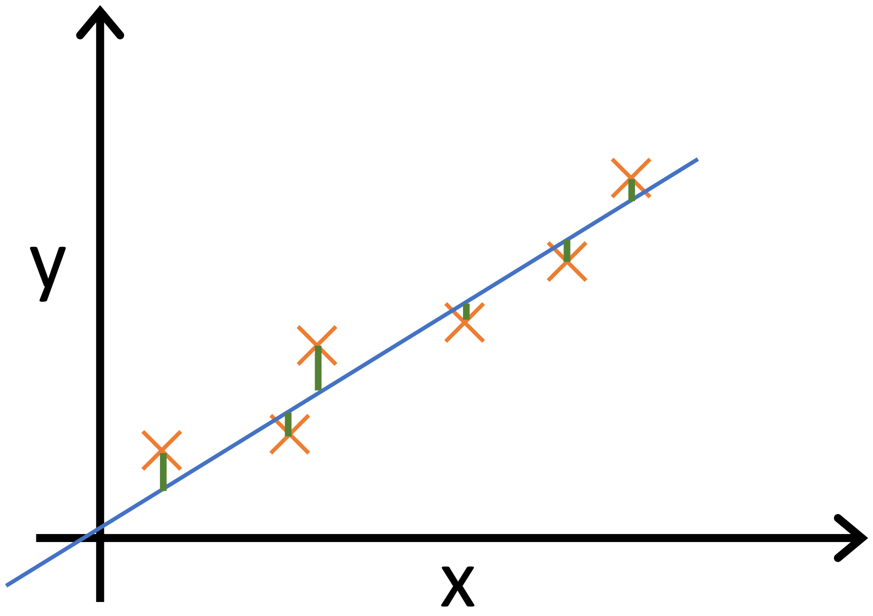
Model 1



Idea: Choose θ_0, θ_1 so that
f(x) is close to y for our
training examples (x, y)

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$ is minimum

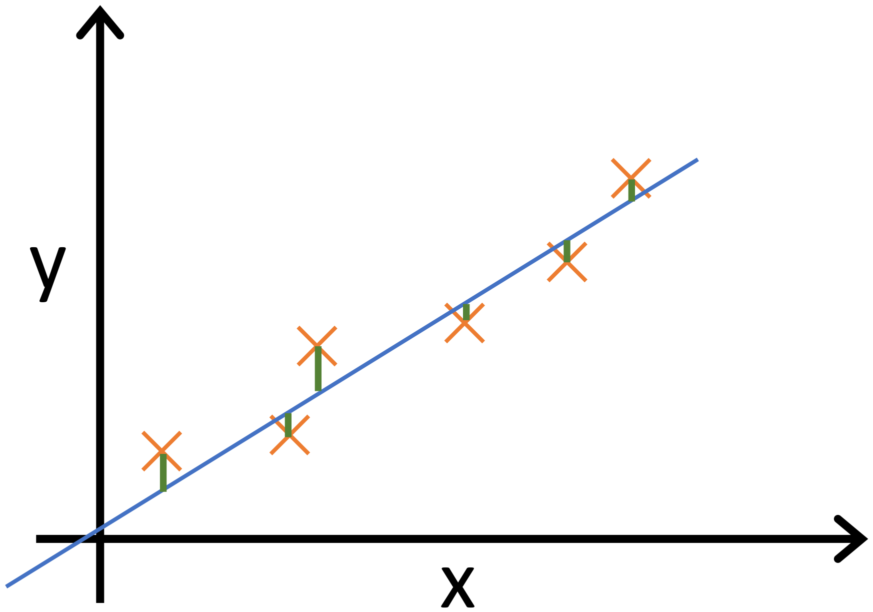
J is the cost function and computes the squared error



How to find θ_0, θ_1 for which J has the minimum value?

$$f(x) = \theta_0 + \theta_1 x$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$



How to find θ_0, θ_1 for which J has the minimum value?

$$f(x) = \theta_0 + \theta_1 x$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$

- J has its minimum where the partial derivates are 0.
- Compute the partial derivatives of J with respect to θ_0, θ_1

$$\frac{\partial}{\partial \theta_0} J(\theta) = \sum_{i=1}^{n}(f(x_i) - y_i)$$

$$\frac{\partial}{\partial \theta_1} J(\theta) = \sum_{i=1}^{n}(f(x_i) - y_i)\, x_i$$

Model 1 – closed form

- By setting the partial derivatives to zero, we obtain a linear system with two equations and two unknowns and can be solved directly to obtain the values for θ_0, θ_1

$$\begin{cases} \theta_1 = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\ \theta_0 = \dfrac{1}{n}\left(\displaystyle\sum_{i=1}^{n} y_i - \theta_1 \sum_{i=1}^{n} x_i\right) \end{cases}$$

- A problem with model 1 is that it cannot represent vertical lines (the slope is undefined with a 0 denominator)

Model 1 – gradient descent

How to find θ_0, θ_1 for which J has the minimum value?

$$f(x) = \theta_0 + \theta_1 x$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$

Solution 2 for minimizing J: Use a more general approach (an iterative optimization algorithm), which is **gradient descent.**

Gradient descent algorithm

Choose initial values for the parameters θ_0, θ_1

- Choose initial value for the learning rate α
- Choose the number iterations
- Convergence is achieved after a fixed number of iterations or when the error (J) is lower than a threshold

$$\text{repeat until convergence } \{$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

$$\}$$

Gradient descent algorithm

$$\text{repeat until convergence } \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$
$$\}$$

Learning rate

Partial derivative

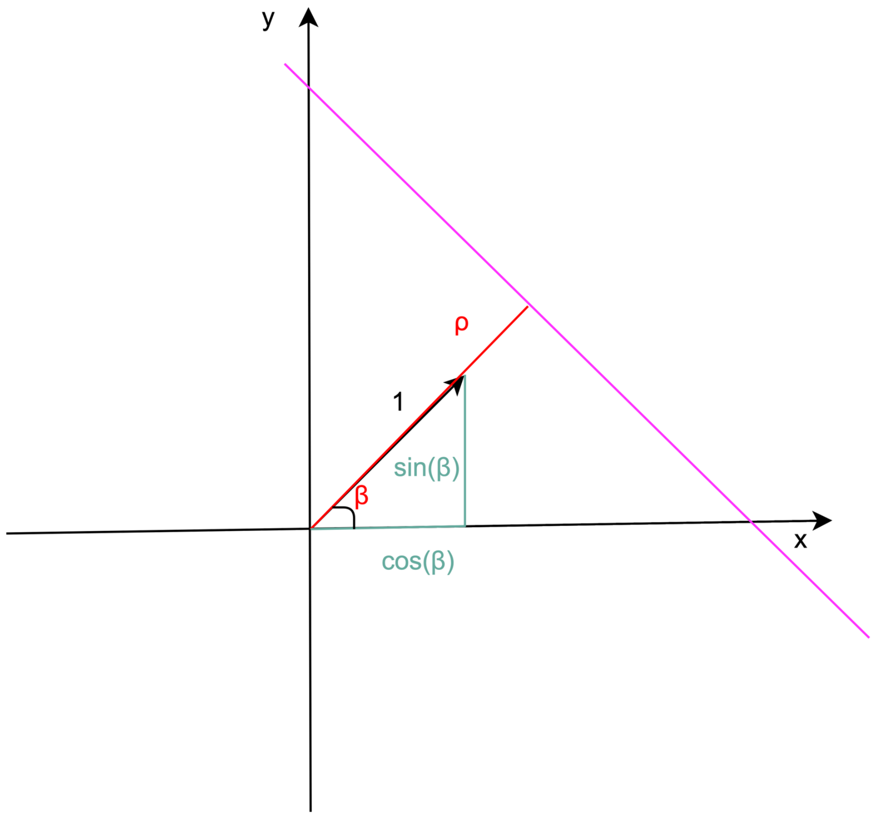
Correct: Simultaneous update

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Incorrect:

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_1 := \text{temp1}$$

Model 2 – Hesse normal form



Equation of a line:

$$x\cos(\beta)+y\sin(\beta)=\rho$$

This describes a line with a unit normal vector of the form [cos(β), sin(β)] that is at a distance ρ from the origin. The parameters of the line are:

- x and y are the Cartesian coordinates of the points on the line.
- ρ is the perpendicular distance from the origin to the line.
- β is the angle between the normal vector of the line (the vector perpendicular to the line) and the positive x-axis.

Model 2

- Equation of a line: **xcos(β)+ysin(β)=ρ**
- The cost function we want to minimize:

$$J(\beta, \rho) = \frac{1}{2}\sum_{i=1}^{n}(x_i cos(\beta) + y_i sin(\beta) - \rho)^2$$

- Compute the partial derivates of J with respect to β and r and set them to 0

$$\frac{\partial J}{\partial \beta} = \sum_{i=1}^{n}(x_i cos(\beta) + y_i sin(\beta) - \rho)\big(-x_i sin(\beta) + y_i cos(\beta)\big)$$

$$\frac{\partial J}{\partial \rho} = -\sum_{i=1}^{n}(x_i cos(\beta) + y_i sin(\beta) - \rho)$$

Model 2

- Closed form solution can be obtained by setting the derivatives to 0

$$\beta = -\frac{1}{2} atan2\left(2\sum_{i=1}^{n} x_i y_i - \frac{2}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i, \sum_{i=1}^{n}(y_i^2 - x_i^2) + \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right)$$

$$\rho = \frac{1}{n}\left(\cos(\beta)\sum_{i=1}^{n} x_i + sin(\beta)\sum_{i=1}^{n} y_i\right)$$

Activitate practica

1. Citiţi datele de intrare din fişierele ataşate. Prima linie conţine numărul de puncte. Liniile următoare conţin perechi (x,y).
2. Afişaţi punctele pe o imagine albă.
3. Utilizaţi modelul 1 si formulele directe pentru a calcula parametri.
4. Utilizaţi modelul 1 si gradient descent pentru a calcula parametri.
5. Folosiţi modelul 2 şi formulele directe pentru a calcula parametri.