AI CAMP

school of ai Algiers

Hello and Welcome to AI Camp

AI CAMP

school of ai Algiers

Linear Regression

Chamel Nadir Bouacha

Workshop Structure

1. Introduction
2. Understanding Linear Regression
3. Types of Linear Regression
4. Methods for Training a Linear Regression Model
5. Evaluation Metrics
6. Conclusion

1- Introduction

AI CAMP

school of ai Algiers

1- Introduction

as we already know , machine learning learning

has two types :

1- Introduction

as we already know , machine learning learning

has two types :

Unsupervised

1- Introduction

as we already know , machine learning learning

has two types :

Unsupervised

Supervised

Supervised Machine Learning

Supervised Machine Learning

1-Regression

AI CAMP

school of ai Algiers

Supervised Machine Learning

1-Regression

2-Classification

Regression

This is when we want to predict a continuous output (aka a number) :

ie: house price prediction in a certain area, medical cost prediction, ...etc

Classification

This is when we want to predict a discrete and finite output (aka 0/1):

ie : prediction if a photo is a dog or no , spam detection model (detect if a mail is a spam or no), ...etc

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

To better understand the Intuition behind Linear regression , we will introduce a famous problem in the world of AI and Machine learning.

2- Understanding Linear Regression

To better understand the Intuition behind Linear regression , we will introduce a famous problem in the world of AI and Machine learning.

House price prediction problem

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

Suppose a client came to you and asked you to predict his house price .and you are provided with the following informations. first we will plot the data to have a better glance of how it looks

Data table

size in $feet^2$	price in $1000's
2104	400
1416	232
1534	315
852	178
...	...
3210	870

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

AI CAMP

school of ai Algiers

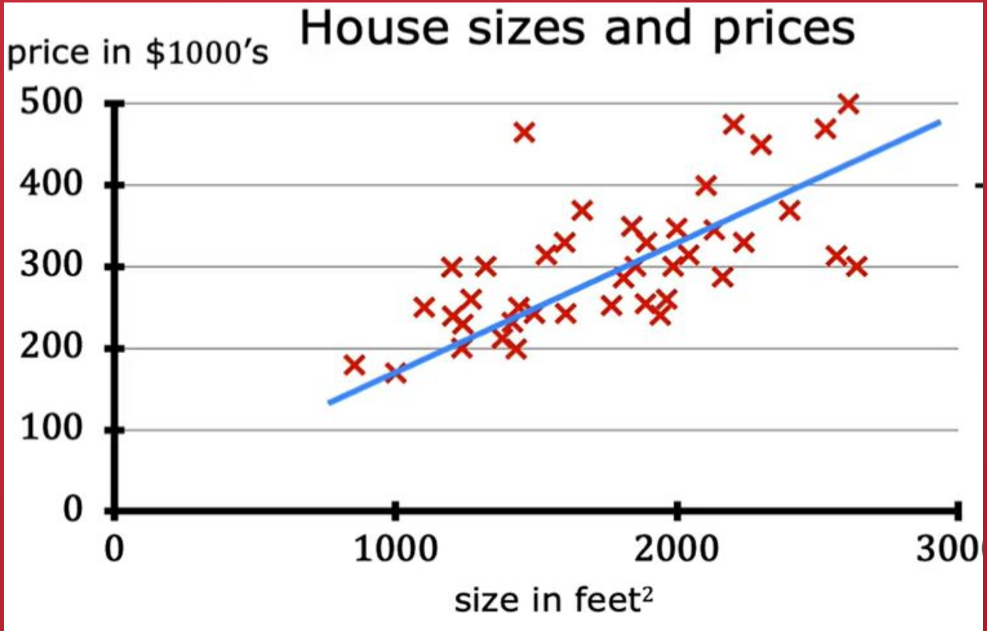
2- Understanding Linear Regression

we create a line that

fits our data and that's

all About

Linear Regression



AI CAMP

school of ai Algiers

AI CAMP

school of ai Algiers

Thank you for attending

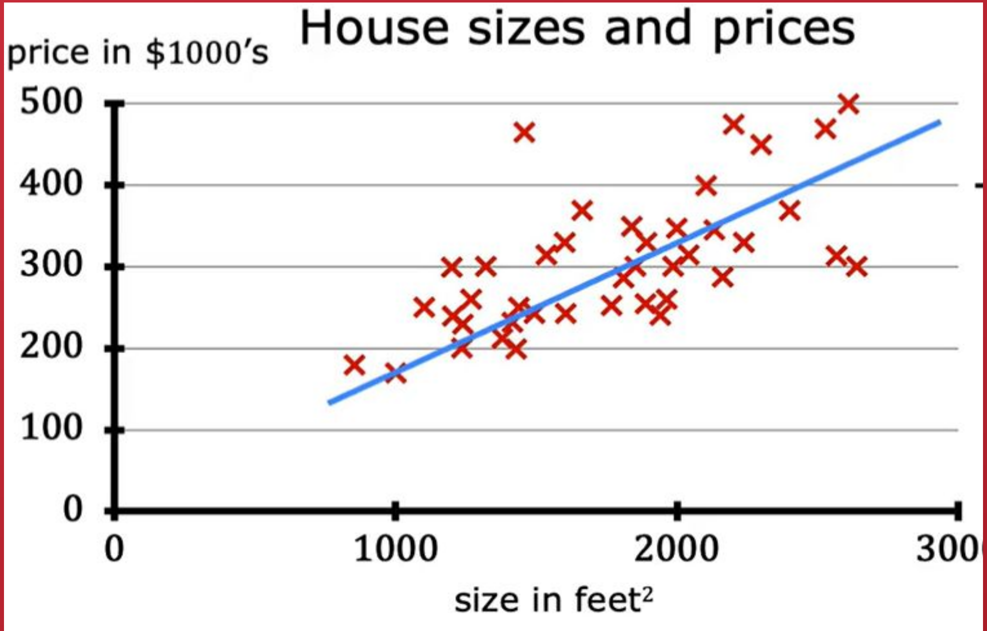
any questions ?

Just Joking LOL

2- Understanding Linear Regression

So here the function in blue has the following equation

$$\hat{y} = \theta_0 + \theta_1 x_1$$



2- Understanding Linear Regression

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

$$\hat{y} = \theta_0 + \theta_1 x_1$$

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

$$\hat{y} = \theta_0 + \theta_1 x_1$$

Our Goal is to extract the values of θ_0 and θ1 that make the function fits well our data !

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

$$\hat{y} = \theta_0 + \theta_1 x_1$$

Our Goal is to extract the values of θ_0 and θ1 that make the function fits well our data !

BUT HOW ?

Notation

- $\hat{y}$ is the predicted value.
- n is the number of features.
- x1 is the feature value.
- θ0 is the bias and the feature weight is θ1
- we define our hθ function as follows

$$\hat{y} = h_{\theta}(x) = \theta \cdot x$$

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

We Should Define then Root Mean Squared Error Cost Function

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

For every row in our dataset we sum :

- we take yi **(in our case it is the house price)** and we subtract from it the value that we predict for the same house ŷi using our linear function
- we square the result .

Data table

size in feet^2	price in \$1000's
2104	400
1416	232
1534	315
852	178
...	...
3210	870

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

Now we should minimize the error of our cost function By finding the right parameters θ0 and θ1 that minimize it .

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

AI CAMP

school of ai Algiers

2- Understanding Linear Regression

In practice, it is simpler to minimize the mean squared error (MSE) than the RMSE, and it leads to the same result (because the value that minimizes a positive function also minimizes its square root).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

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

3- Types Of Linear Regression

Let's continue from our previous example of house price prediction , we had only 1 feature (size in $feet^2$) .

This is called simple Linear regression or univariate linear regression

3- Types Of Linear Regression

Suppose now we don’t only have size as feature but also the year it was built in , and the ocean proximity

This is called Multiple Linear regression or Multivariate linear regression

AI CAMP

school of ai Algiers

3- Types Of Linear Regression

Most of the linear regression problems are multivariate linear regression , since using multiple features will give you more information about your dataset so the performance of the model will be better

3- Types Of Linear Regression

So now the equation of ŷ will become like that

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

With $\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta} \cdot \mathbf{x}$

4- Methods For Training a Linear Regression Model

4- Methods For Training a Linear Regression Model

There are many ways to minimize the cost function we defined earlier , we will only mention 2 of them in this workshop

4- Methods For Training a Linear Regression Model

There are many ways to minimize the cost function we defined earlier , we will only mention 2 of them in this workshop

The Normal Equation

4- Methods For Training a Linear Regression Model

There are many ways to minimize the cost function we defined earlier , we will only mention 2 of them in this workshop

The Normal Equation

Gradient Decent

4- Methods For Training a Linear Regression Model

The normal Equation method is an analytical approach used to find the exact parameters (θ) that minimize the cost function

It has the following formula

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{y}$$

The Normal Equation method

Benefits

- It provides a closed-form solution, meaning it solves the problem in one calculation.
- The Normal Equation is computationally efficient when the number of features (n) is small

Inconvenients

- However, if the number of features is large, the computational complexity increases, as it involves matrix inversion.

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}^\intercal \mathbf{X}\right)^{-1} \mathbf{X}^\intercal \mathbf{y}$$

AI CAMP

school of ai Algiers

What’s the Alternative then ?

Gradient descent

Let's go back to our first example :

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\; J(\theta_0, \theta_1)$

Gradient descent

We will use the following for each parameter θ_j until we reach the value that minimize the previous cost function

$$\text{Repeat until convergence } \{$$

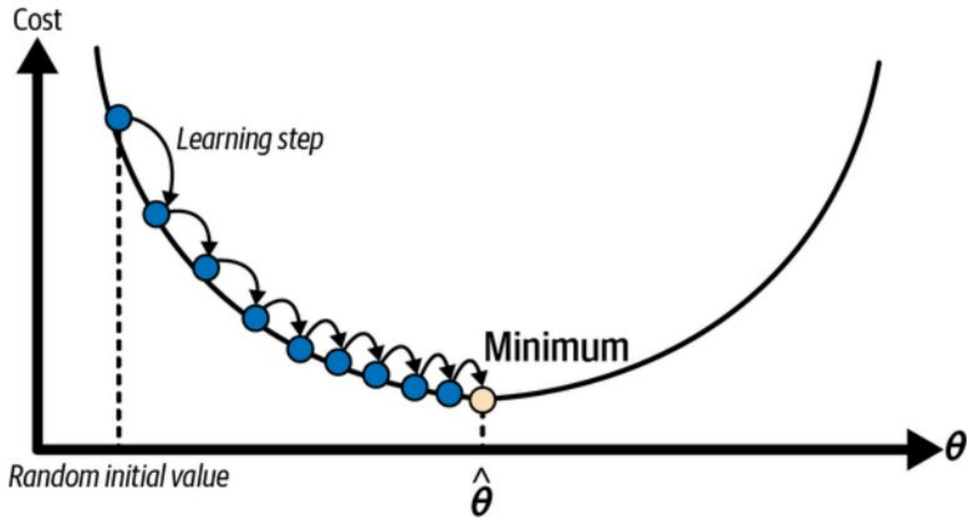
$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\}$$

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

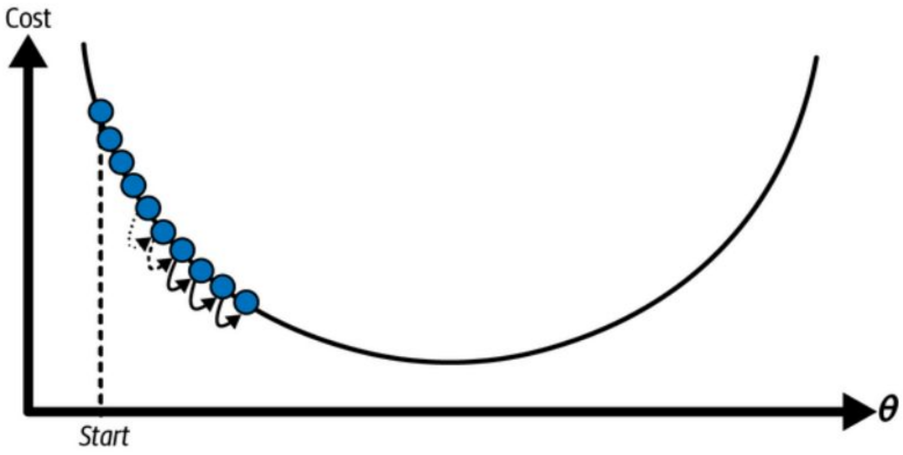
Gradient descent

With the right value of the Learning Rate (Learning Step) we will have a learning curve that looks just like that



Gradient descent

If the learning rate is too small, then the algorithm will have to go through many iterations to converge, which will take a long time

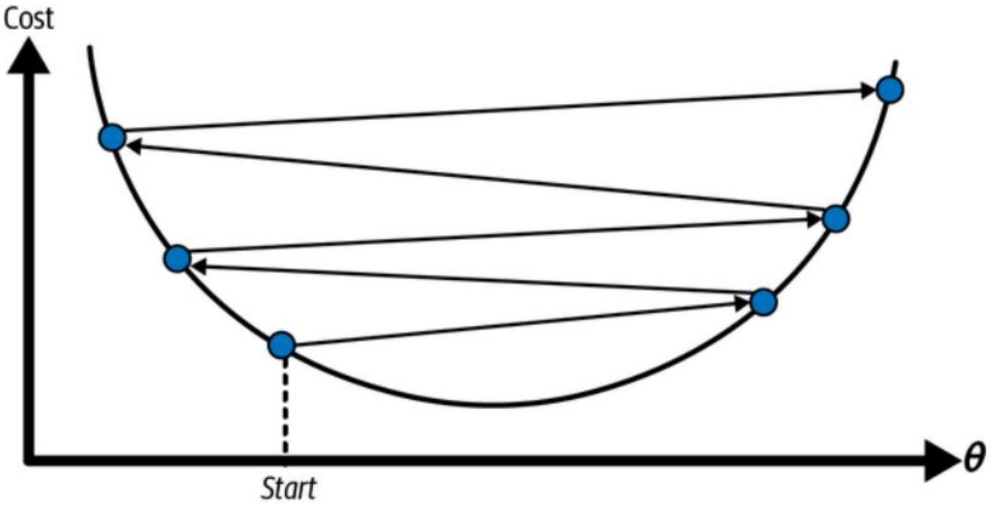


AI CAMP

school of ai Algiers

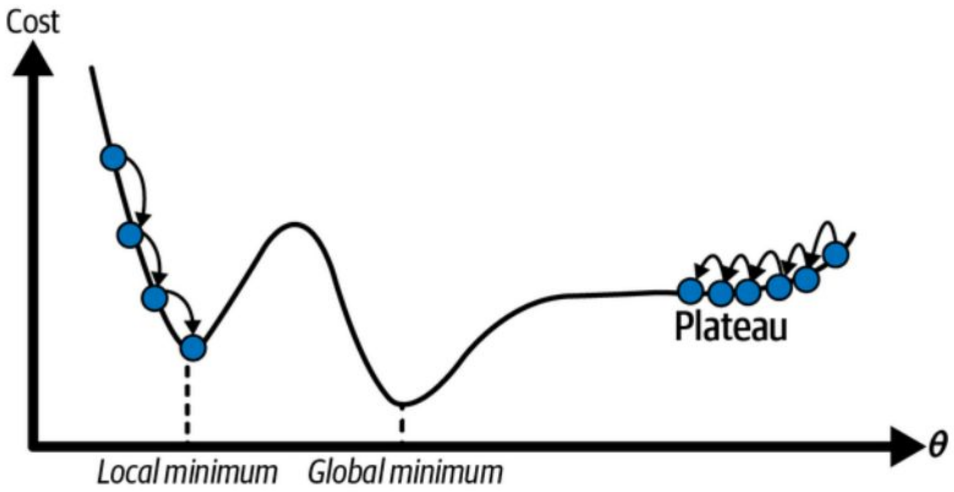
Gradient descent

On the other hand, if the learning rate is too high, you might jump across the valley and end up on the other side, possibly even higher up than you were before



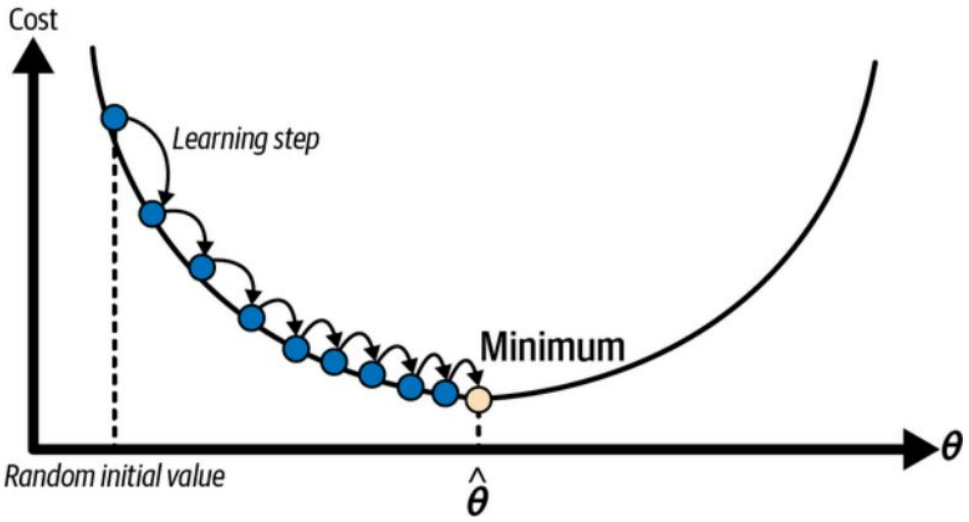
Gradient descent

Additionally, not all cost functions look like nice, regular bowls. There may be holes, ridges, plateaus, and all sorts of irregular terrain, making convergence to the minimum difficult.



Gradient descent

Fortunately, the MSE cost function for a linear regression model happens to be a convex function. so we will always have a learning curve that looks similar to this



AI CAMP

school of ai Algiers

4- Methods For Training a Linear Regression Model

1) **Data Preprocessing:**
   Imagine you're preparing ingredients for a recipe. Data preprocessing is like washing, chopping, and organizing those ingredients before you cook.

2) **Handling Missing Values:**
   Some ingredients might be missing (data missing). Decide if you can still cook with what you have or if you need to replace missing ingredients.

3) **Handling Outliers:**
   Outliers are like unexpected ingredients – they can throw off your recipe. Decide if you keep them or replace them with something more common.

4) **Feature Scaling:**
   Feature scaling is like making sure all ingredients are in the same units. If one ingredient is in grams and another in kilograms, the recipe might not turn out well

5- Evaluation Metrics

We always hear people bragging about their model's accuracy.

So, how do we calculate our Linear Regression model's accuracy?

Long story short : we can't !

Accuracy is an evaluation metric for classification tasks.

How do we **evaluate** our model then ?

5- Evaluation Metrics

- **Mean Squared Error (MSE).**

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

- **R-squared Score (R2).**

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

- **Mean Absolute Error (MAE)**

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Time To practice !

let’s head to our notebooks

AI CAMP

school of ai Algiers

Thank you for attending

any questions ?

Thanks for iness that contributed to the workshop preparation