Predicting Relationships Between Variables with Regression

Swetha Kolalapudi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Recognize Regression problems in different fields : from Quant Trading to Demand forecasting

Understand how to set up a Regression problem - Dependent and Independent variables

Contrast Classification and Regression and when each technique should be used

Examples of Regression Problems

What will be the returns from a stock on a given date?

If waiting time increases, how does this affect customer satisfaction?

What will be the sales of this product in a given week?

A Regression Problem Statement

You would like to compute some Continuous Value

Stock Returns

Sales

A Regression Problem Statement

You would like to quantify the relationship between 2 variables

Wait time  **Customer Satisfaction**

Regression

Regression helps you quantify the relationship between variables

Let's take the example of Stock Returns

Day of the week

Day of the month

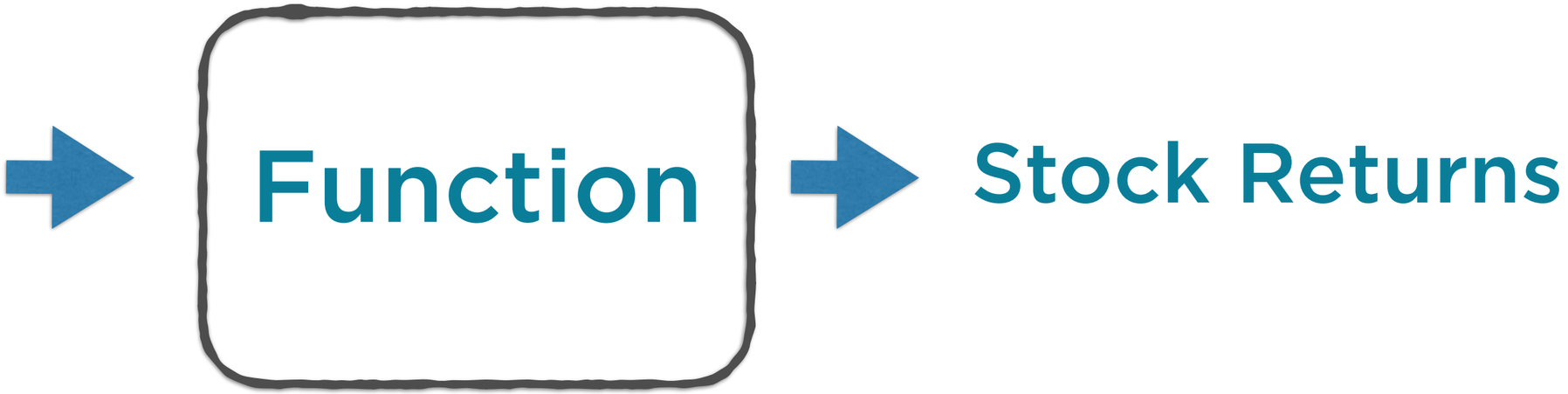
Daily Stock returns in the last 1 week

Stock Returns for a given day might depend on a number of variables

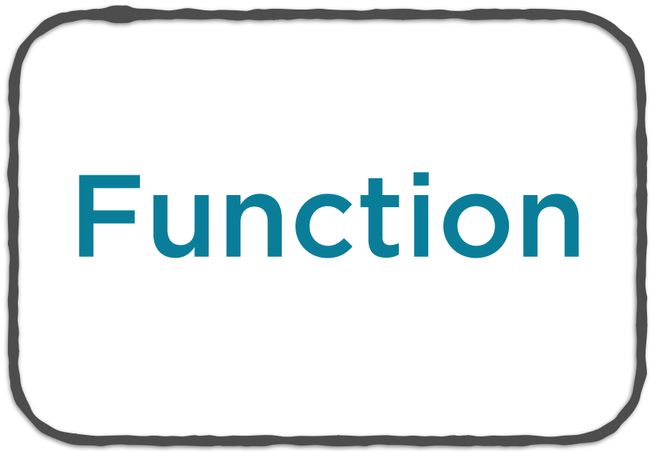
Day of the week

Day of the month

Daily Stock returns in the last 1 week

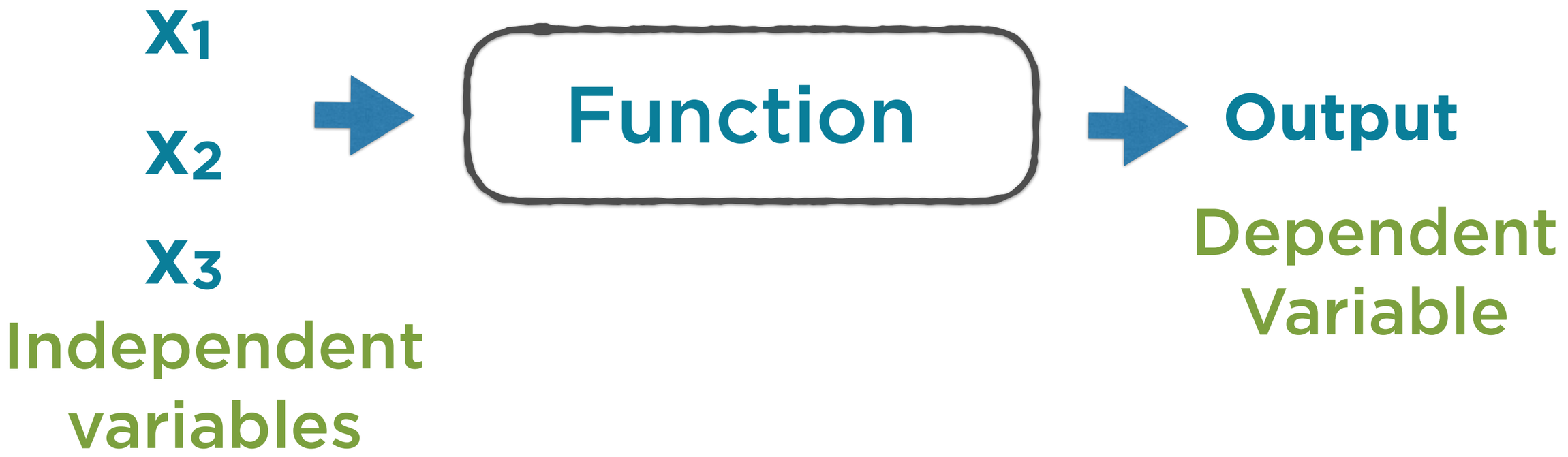


The Stock Returns on a given day are a function of these variables



Regression is the process that identifies this function

Regression


X_1
X_2
X_3
Independent variables
Function
Output
Dependent Variable

Types of Regression

Linear Regression

Polynomial Regression

Non-linear Regression

Demand Forecasting

Many businesses need to estimate sales at a future point in time

Demand Forecasting

A retail business will use it for inventory planning

Demand Forecasting

A manufacturing business will use it plan production cycles

Demand Forecasting

Sales at a future time might depend on

Sales in the previous week

Expected Marketing spend

Holidays

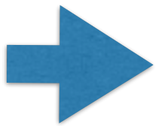
Regression

Sales in the previous week

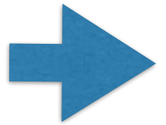
Expected Marketing spend

Holidays

Independent variables



Function



Sales in future week

Dependent Variable

High Risk, High Reward

A maxim heard quite often in the business world!

The Capital Asset Pricing Model

Used for pricing risky securities

$$R_i = R_f + \beta_i (R_m - R_f)$$

The CAPM Model

$$R_i = R_f + \beta_i (R_m - R_f)$$

The return on a security

The CAPM Model

$$R_i = R_f + \beta_i (R_m - R_f)$$

The risk free Rate of Return

+ a Premium

The CAPM Model

$$R_i = R_f + \beta_j (R_m - R_f)$$

The Volatility of the security

A Measure of risk

The CAPM Model

$$R_i = R_f + \beta_i (R_m - R_f)$$

Expected return of the market over and above the risk free rate

The CAPM Model

$$R_i = R_f + \beta_j (R_m - R_f)$$

Beta of a security is an important measure

It can tell you how much risk a security adds to a portfolio

The CAPM Model

$$R_i = R_f + \beta_j (R_m - R_f)$$

Beta quantifies the relationship between Stock Returns and Market Returns

The CAPM Model

$$R_i = R_f + \beta_j (R_m - R_f)$$

Use Regression to find the value of Beta

Facial Feature Detection

Teach a computer to identify eyes, nose, mouth

Facial Feature Detection

Facial Recognition

Virtual dressing rooms

Auto-capture photos

Facial Feature Detection

Find the co-ordinates of the important facial features

Facial Feature Detection

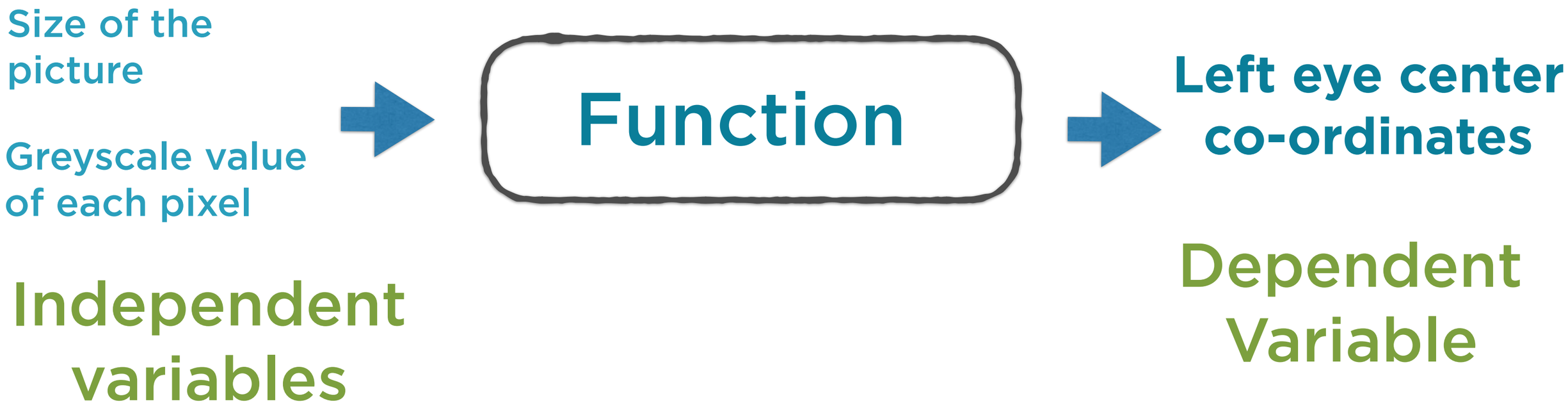
Their position depends on

- relative position within the picture
- the properties of surrounding pixels

Facial Feature Detection

The co-ordinates of each feature can be found using one regression problem

Regression


Size of the picture
Greyscale value of each pixel
Independent variables
Function
Left eye center co-ordinates
Dependent Variable

Classification and
Regression are
similar in many ways

Typical Classification Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Problem Statement

The problem statement in Classification

Assign a category/label

The problem statement in Regression

Compute a continuous value

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

"Train a model" using the training data

Test

"Test the model" using test data

Features

In classification

The problem instance represented using numeric attributes

In Regression

The Independent Variables

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

"Train a model" using the training data

Test

"Test the model" using test data

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Training

In classification

Use training data to build a classifier

In Regression

Use training data to quantify the relationship between variables

Training

Regression

is also a form of Supervised learning

Typical Regression Setup

Problem Statement

Define the problem statement

Features

Represent the training data and test data using numerical attributes

Training

“Train a model” using the training data

Test

“Test the model” using test data

Typical Regression Setup

Problem Statement	Features	Training	Test
Define the problem statement	Represent the training data and test data using numerical attributes	"Train a model" using the training data	**"Test the model" using test data**

Test

In classification

Assign a label for a new instance

In Regression

Compute output value given the independent variables

Classification vs Regression

A Categorical output

A Continuous output

Classification vs Regression

Relationship between Input and Output

Usually a black box

A mathematical function

Summary

Recognize Regression problems in different fields : from Quant Trading to Demand forecasting

Understand how to set up a Regression problem - Dependent and Independent variables

Contrast Classification and Regression and when each technique should be used