Member-only story

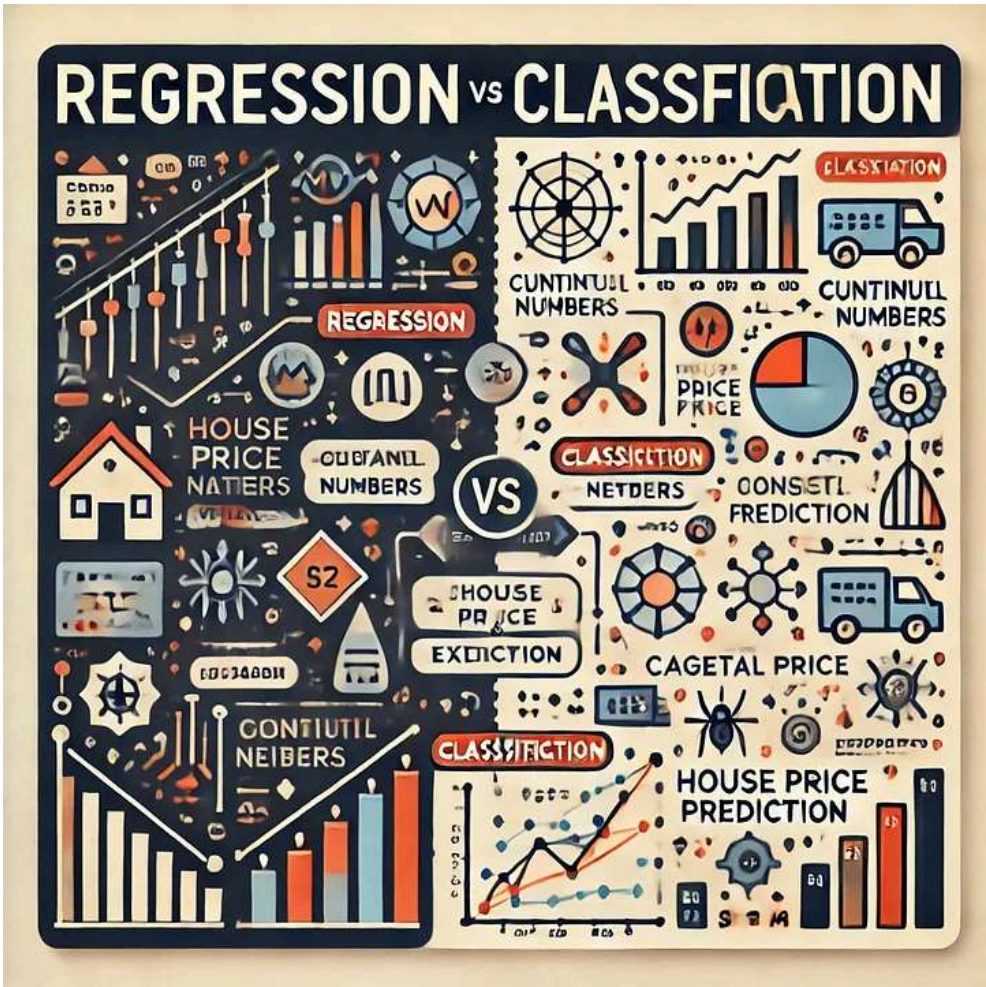
Regression vs. Classification: How to Choose the Right Machine Learning Approach

Ritesh Gupta · Follow
5 min read · 3 days ago

A Comprehensive Guide to Understanding When and Why to Use Regression or Classification Algorithms

Regression vs. Classification: Which One to Use?

Algorithm choice is critical in machine learning, especially for prediction and insight. Of the most commonly used algorithms are regression and classification. Both of these supervised learning techniques are different in purpose, application, and method for achieving their goals. In this article, we discuss key differences between regression and classification, when to use them, and look at some real-world examples to clarify these concepts.



1. Understanding Supervised Learning

Understanding Both Regression and Classification are the Types of Supervised Learning wherein the model is learned in an environment with a labelled set of data. It comprises data of the type consisting of the input features and their output labels and enables the model to map inputs to outputs. Therefore, the only difference between the two is the kind of output.

- **Regression** is used for predicting a **continuous output** (e.g., predicting house prices).
- **Classification** is used for predicting a **discrete output** (e.g., classifying emails as spam or not spam).

2. What is Regression?

Regression is a kind of predictive modeling where the variable to be estimated is of continuous nature. This process helps predict a range value based on previously taken measurements. These variables include any

prices, temperatures, and sometimes even sales, along with some time variables.

Linear Regression — Output variable is projected depending on input variables, which are assumed to lie in a straight line.

Polynomial Regression — Can be undertaken when relationship between variable is nonlinear

Logistic Regression — This is although called as regression but taken for binary classification.

Regression Example:

Let's assume you want to predict house price depending upon features like square feet, location, and so forth for bedroom space. Since house price values are continuous, in that case, linear regression should be your algorithm of choice for modeling such data. You may take any model trained on historical prices, and it can generate some reasonable price for any new data based on its patterns learned from the existing set of data.

Key Features of Regression:

Output Type: Continuous (for example, numbers that may be decimal or integer).

Use Case: For those problems where precise values are needed, for example, in the prediction of sales revenue, stock prices, etc.

Evaluation Metrics: MSE, RMSE, MAE, R-Squared.

3. What is Classification?

Classification is used for predictive modeling when the target variable is categorical. In classification, we predict which category or class an observation belongs to, given the input features.

Types of Classification:

Binary Classification — Only two classes possible (for example spam or not spam).

Multiclass Classification — More than two classes (like identifying pictures as cats, dogs, or birds). Multilabel Classification-Class instances can be multiple, that is, instances can be members of two or more classes at the same time (for example, a label assigned to an article as "technology" and "science").

Classification Example:

It'd help if you were designing an email spam identification system: you would like to be able to classify each into either "spam" or not spam" based on whether the content, sender, and subject lines have something in common. In these cases, binary classification works well. Once the model was trained on a labeled collection of emails, it continues to classify new emails following the patterns learned.

Class Characteristics:

Output Type- Categorical (e.g., labels, classes). Use Case: For type of problems that are categorical or classifying fraud type, medical diagnosis, etc.

Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, Area Under the Curve (AUC).

4. Regression vs. Classification: Key Differences

Aspect	Regression	Classification
Output	Continuous values (e.g., prices)	Categorical values (e.g., classes)
Goal	Predict exact numerical values	Predict categories or labels
Algorithms	Linear Regression, Polynomial Regression	Logistic Regression, Decision Trees, SVM, etc.
Metrics	MSE, RMSE, MAE, R-Squared	Accuracy, Precision, Recall, F1-Score
Examples	Predicting house prices, stock prices	Email spam detection, image recognition

---

5. When to Use Regression vs. Classification

Use Regression When:

Your target variable is continuous and can accept any value in some range. You want to predict an exact value (for example, sales figures, temperature forecast).

Examples: Predict customer lifetime value, Stock Price Forecasting, Demand forecasting.

Use Classification When:

The target variable is categorical, having discrete classes. It has to classify instances into pre-determined classes-for example, whether someone says one of the things listed in your opinion columns is the future of online news (for example, it could be, " Sentiment Analysis"; it might be risk classification). Sentiment classification (social media analysis); Image classification; Loan status to approve.

6. Real-World Examples: Regression vs. Classification

Case 1: Customer churn

Problem Statement: Telco want to forecast which customer would churn and will not. Solution: This is a binary classification problem because the output is a choice between two classes: "churn" or "not churn."

Case 2: House Price Prediction

Problem: A real estate agency wants to predict the sale price of a house based on features like size, location, and age. Solution: This is a regression problem because the target variable, house price, is continuous.

Case 3: Classifying Disease Type in Medical Diagnosis

Problem: Based on symptoms and medical history, a hospital wants to classify a disease as belonging to one of multiple classes of diseases (flu, cold, COVID-19, etc). Solution: This is a multi-class classification problem since the classes are more than two.

Case 4: Predicting Future Sales Volume

Problem: A retail firm wants to predict units to sell next month Solution: This is regression since the sales volume will be continuous.

7. Choosing the Right Approach: Practical Tips

To choose between regression and classification:

Identify the type of target variable you are working with: continuous or categorical? Look at your data set: will your data be more ideal for specific-

value prediction or category-based decisions?

Business objectives: do you want an exact prediction for a specific value (regression) or category-based determination (classification)?

Measure evaluation metrics: if you need more precision regarding absolute error measurements, use regression measures — MSE or RMSE.

8. Conclusion

In consequence, there is, accordingly, the need to know this key difference between regression and classification for picking out an appropriate ML model approach. You can do the best regression for continuous value predictions using the right classification where varying kinds of data fall in specific classes. In doing that way, one will figure out an appropriate algorithm through which one can do with the models that have an ability to provide valuable insights and relevant prediction accordingly on the basis of different sources of data. Whether you are trying to predict the price of stocks or classify customer reviews, knowing when to use regression versus classification will increase your accuracy and utility of machine learning models.

Data Science | Machine Learning | Deep Learning | Python | NLP

Written by Ritesh Gupta

Follow

3.5K Followers

Data Scientist, I write Article on Machine Learning| Deep Learning| NLP | Open CV | AI Lover ❤️

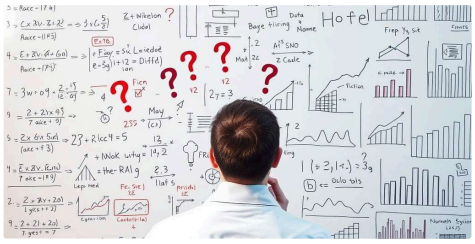
More from Ritesh Gupta

Ritesh Gupta

10 Must-Try LLM Projects to Boost Your Machine Learning Portfolio

🚀 10 Exciting LLM Projects to Elevate Your Machine Learning Skills! 🧠

Oct 6 51 1



Ritesh Gupta

Can You Handle These 25 Toughest Data Science Interview Questions?

The role of a Data Scientist demands a unique blend of skills, including statistics, machine...

Sep 25 62

Ritesh Gupta

14 Life Changing Lessons From Chanakya Niti everyone should...

Who is Chanakya?

Jan 30, 2023 110 1

Ritesh Gup... in Artificial Intelligence in Plain Engli...

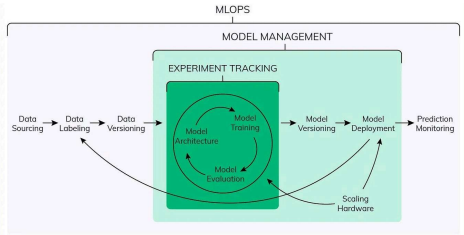
From Jupyter to Production: Deploying Machine Learning...

Turn your Jupyter Notebook experiments into production-ready applications with this...

Oct 24 10

See all from Ritesh Gupta

Recommended from Medium

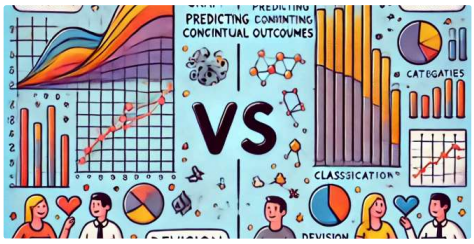


Prashant Shinde in Data And Beyond

From Zero to MLOps Hero: Your First Steps in Building an MLOps...

Just a few months ago, I was where you are—thrilled by the possibilities of machine...

Sep 16 110



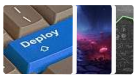
Vikash Singh

Regression vs. Classification: the Davengers Style

Beginner friendly introduction in a fun-friendly manner!

Oct 24 36

Lists

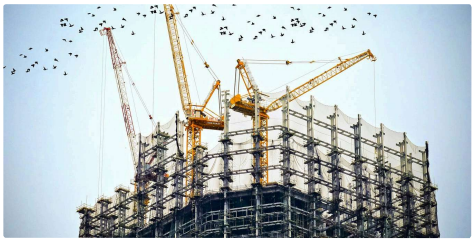


Predictive Modeling w/ Python
20 stories · 1638 saves

Practical Guides to Machine Learning
10 stories · 2005 saves

Natural Language Processing
1793 stories · 1407 saves

Coding & Development
11 stories · 889 saves

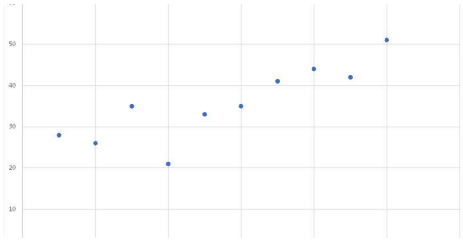


Leonardo Anello in Towards Data Science

Practical Guide to Data Analysis and Preprocessing

Techniques for data cleaning, transformation, and validation to ensure quality data

6d ago 135

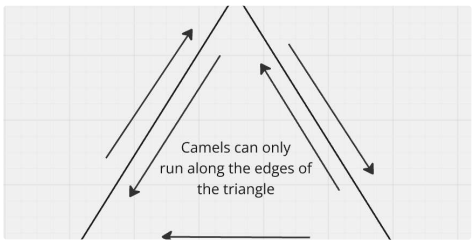


BlogInnovazione

Machine Learning example with Python: Simple Linear Regression

In this machine learning example we are going to see a linear regression with only on...

Aug 3 21



 Lucas Samba

3 Probability Questions I was asked in Walmart Data Scientist Interview

Recently I got an opportunity to interview at Walmart for Data Scientist—3 position. All...

Aug 23 237 4

 Abdur Rahman in Stackademic

Python is No More The King of Data Science

5 Reasons Why Python is Losing Its Crown

Oct 23 3K 19

See more recommendations