Ordinal Logistic Regression

An overview and implementation in R

Akanksha Rawat · Follow

Published in Towards Data Science

5 min read · Feb 20, 2018

Listen Share More

Fig 1: Performance of an individual — Poor, Fair, Excellent

Can you guess what is the common link in the variables mentioned below:

- Job satisfaction level — *Dissatisfied, Satisfied, Highly Satisfied*
- Performance of an individual — *Poor, Fair, Excellent*
- Impact of a regulation on bank's performance — *Positive, Neutral, Negative*

The variables are not only categorical but they are also following an order (low to high / high to low).

If we want to predict such **multi-class ordered variables** then we can use the ***proportional odds logistic regression*** technique.

Objective

To understand the working of Ordered Logistic Regression, we'll consider a study from World Values Surveys, which looks at factors that influence people's perception of the government's efforts to reduce poverty.

Our objective is to predict an individual's perception about government's effort to reduce poverty based on factors like individual's country, gender, age etc. In the given case study, individual's perception can take the following three values - *Too Little, About Right, Too Much.*

For our analysis, we'll be using data from the World Values Surveys for Australia, Norway, Sweden, and the United States from 'carData' package in R.

```
library(carData)
library(MASS)

data(WVS)
head(WVS)
```

```
> head(WVS)
       poverty religion degree country age gender
1   Too Little      yes     no     USA  44   male
2  About Right      yes     no     USA  40 female
3   Too Little      yes     no     USA  36 female
4     Too Much      yes    yes     USA  25 female
5   Too Little      yes    yes     USA  39   male
6  About Right      yes     no     USA  80 female
>
```

Fig 2 — Dataset

Description of the data

Poverty is the multi-class ordered dependent variable with categories — 'Too Little', 'About Right' and 'Too Much'. We have the following five independent variables

- **Religion:** member of a religion -no or yes
- **Degree:** held a university degree -no or yes
- **Country**: Australia, Norway, Sweden or the USA
- **Age:** age (years)
- **Gender**: male or female

Let's now analyze the **descriptive statistics** for this dataset:

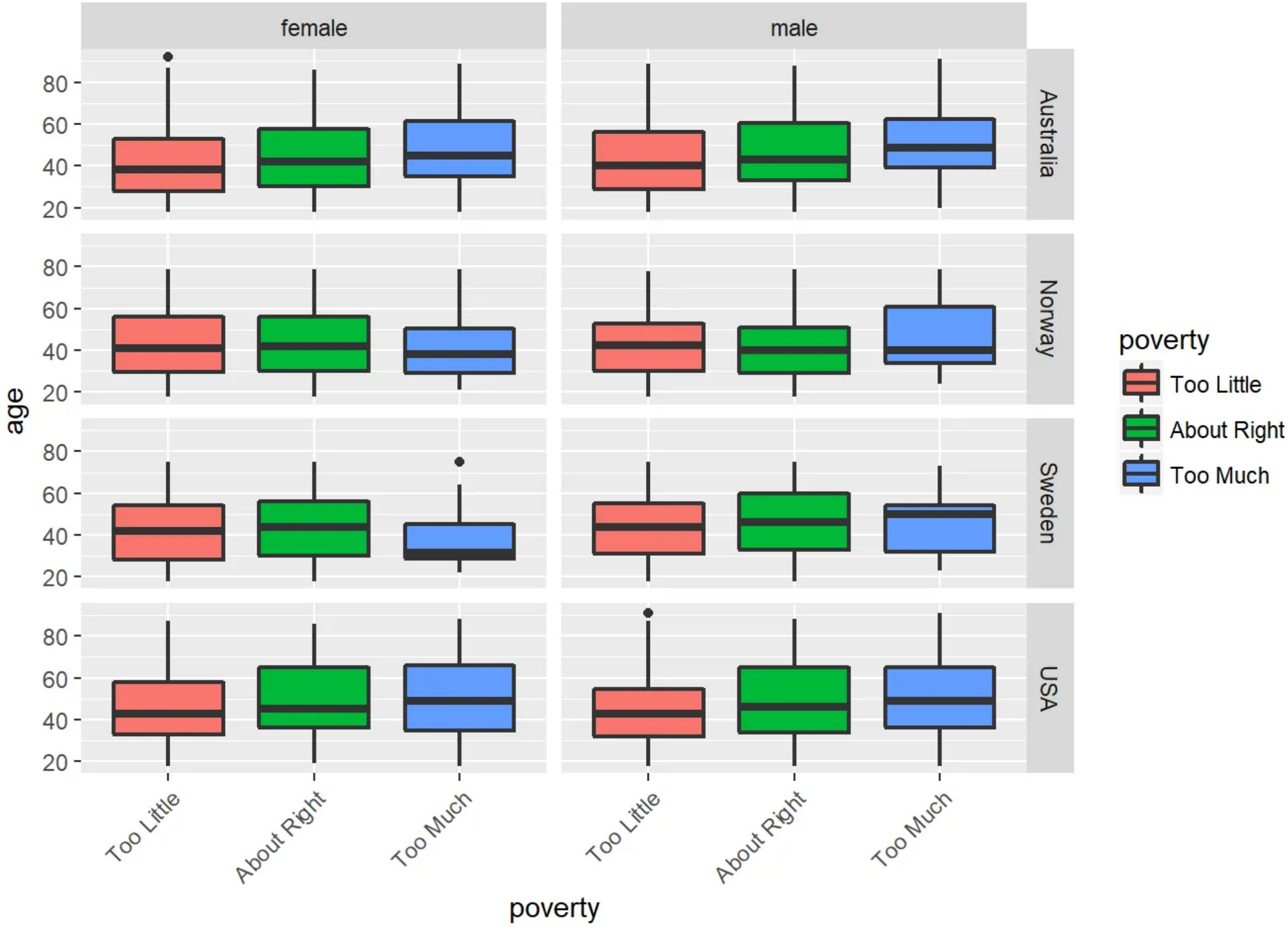
```
summary(WVS)
```

```
> summary(WVS)
        poverty      religion    degree           country          age           gender
 Too Little :2708   no : 786   no :4238   Australia:1874   Min.   :18.00   female:2725
 About Right:1862   yes:4595   yes:1143   Norway   :1127   1st Qu.:31.00   male  :2656
 Too Much   : 811                         Sweden   :1003   Median :43.00
                                          USA      :1377   Mean   :45.04
                                                           3rd Qu.:58.00
                                                           Max.   :92.00
```

Fig 3 — Descriptive Statistics

We can also analyze the distribution of poverty across age, sex and country

```
ggplot(WVS, aes(x = poverty, y = age, fill = poverty)) +
geom_boxplot(size = .75) +   facet_grid(country ~ gender, margins =
FALSE) +   theme(axis.text.x = element_text(angle = 45, hjust = 1,
vjust = 1))
```



Fitting the Model

We'll now fit the Proportional Odds Logistic Regression model using **polr** function from the **MASS** package.

```
model_fit <- polr(poverty~religion+degree+country+age+gender, data =
WVS, Hess = TRUE)
summary(model_fit)
```

```
> summary(model_fit)
Call:
polr(formula = poverty ~ religion + degree + country + age +
    gender, data = WVS, Hess = TRUE)

Coefficients:
                  Value Std. Error t value
religionyes     0.17973   0.077346   2.324
degreeyes       0.14092   0.066193   2.129
countryNorway -0.32235   0.073766  -4.370
countrySweden -0.60330   0.079494  -7.589
countryUSA      0.61777   0.070665   8.742
age             0.01114   0.001561   7.139
gendermale      0.17637   0.052972   3.329

Intercepts:
                          Value   Std. Error t value
Too Little|About Right   0.7298  0.1041      7.0128
About Right|Too Much     2.5325  0.1103     22.9496

Residual Deviance: 10402.59
AIC: 10420.59
>
```

Fig 4 — Model Summary

In the output above, we get the information about

- Model equation
- The regression coefficients with their values, standard errors and t value. There is no significance test by default but we can calculate p-value by comparing t value against the standard normal distribution.
- Estimates for two intercepts
- Residual deviance and AIC, which are used in comparing the performance of different models

The significance of coefficients and intercepts

```
summary_table <- coef(summary(model_fit))
pval <- pnorm(abs(summary_table[, "t value"]),lower.tail = FALSE)* 2
summary_table <- cbind(summary_table, "p value" = round(pval,3))
summary_table
```

```
> summary_table
                           Value  Std. Error   t value p value
religionyes           0.17973194 0.077346042  2.323738   0.020
degreeyes             0.14091745 0.066193109  2.128884   0.033
countryNorway        -0.32235359 0.073766034 -4.369946   0.000
countrySweden        -0.60329785 0.079493909 -7.589234   0.000
countryUSA            0.61777260 0.070664761  8.742301   0.000
age                   0.01114091 0.001560585  7.138933   0.000
gendermale            0.17636863 0.052972253  3.329453   0.001
Too Little|About Right 0.72976353 0.104061643  7.012800  0.000
About Right|Too Much  2.53247870 0.110349780 22.949558   0.000
```

Fig 5 — Model Significance

Since the p-value for all the variables <0.05, hence they are statistically significant at 95% CI

Interpretation of the Proportional Odds Model

In order to interpret this model, we first need to understand the working of the proportional odds model.

Let J be the total number of categories of the dependent variable and M be the number of independent variables (In the given dataset, J=3 and M = 5).

The mathematical formulation of the Proportional Odds Model is given below

$$\mathit{logit}\,[P(Y \leq j)] = \alpha_j - \sum \beta_i X_i$$

$$\text{where } j = 1,\ldots,J-1 \text{ and } i = 1,\ldots,M$$

Fig 6 — Equation for Proportional Odds Model

Here, j is the level of an ordered category with J levels and i corresponds to independent variables

In our case

- j = 1 refers to 'Too Little'
- j = 2 refers to 'About Right'
- j = 3 refers to 'Too Much'

- i = 1 refers to 'religion'
- i = 2 refers to 'degree'

Interpreting the model statistics in Fig 4

Coefficients :

- The categorical variables like **gender** can be interpreted as: a male individual, as opposed to a female individual, is associated with a higher likelihood of having a positive perception about government's efforts to reduce poverty. The t-value is greater than 2 and therefore is statistically significant at the 5% level.
- The continuous variables like **age** can be interpreted as : with one unit increase in age the log of odds of having a positive perception about government's efforts to reduce poverty increases by 0.011

Intercepts:

- Mathematically, the intercept '**Too Little | About Right**' corresponds to **logit[P(Y ≤ 1)].** It can be interpreted as the log of odds of believing that the government is doing 'Too Little' versus believing that the government is doing 'About Right' or 'Too Much'
- Similarly, the intercept '**About Right | Too Much**' corresponds to **logit[P(Y ≤ 2)].** It can be interpreted as the log of odds of believing that the government is doing 'Too Little' or 'About Right' versus believing that the government is doing 'Too Much'

Making predictions on new data

Let say we want to predict the probability corresponding to each perception for an individual — *Test_Person* with the following characteristics

- Religion : yes
- Degree : no
- Country : Norway

- Age : 30
- Gender : male

1. *Mathematical Computation*

By using the intercept and slope values from the Model Summary, we can estimate the desired probabilities in the following manner

The probability corresponding to **Too Little** perception will be calculated as:

logit[P(Y ≤ 1)] = 0.7298 -[(0.17973*1)+(0.14092*0)+(-0.32235*1)+(0.01114*30)+(0.17637*1)]

=> logit[P(Y ≤ 1)] =0.36185

=> P(Y ≤ 1)= exp(0.36185)/(1+exp(0.36185)) = 0.589

In our case, **P(Y ≤ 1) = P(Y =1) = 0.589**

Similarly, the probability corresponding to **About Right** perception will be calculated as:

logit[P(Y ≤ 2)] = 2.5325 -[(0.17973*1)+(0.14092*0)+(-0.32235*1)+(0.01114*30)+(0.17637*1)]

=> logit[P(Y ≤ 2)] =2.16455

=> P(Y ≤ 2)= exp(2.16455)/(1+exp(2.16455)) = 0.897

Hence, P(Y = 2) = P(Y ≤ 2) — P(Y ≤ 1) = 0.897 -0.589

=> **P(Y = 2) = 0.308**

The probability corresponding to **Too Much** perception will be calculated as:

Thus, P(Y = 3) = 1-P(Y ≤2)

=> P(Y = 3) = 0.103

2. Computation in R

Fortunately, we can bypass the above mathematical calculation by using the **predict** function in R

```
new_data <- data.frame("religion"=
"yes","degree"="no","country"="Norway","age"=30,"gender"="male")

round(predict(model_fit,new_data,type = "p"), 3)
```

```
> round(predict(model_fit,new_data,type = "p"), 3)
 Too Little About Right    Too Much
      0.589        0.308       0.103
>
```

Fig 7 —Model Prediction

Our model predicts that the individual *Test_Person* believes that the government's effort to reduce poverty are *Too Little*

If you wish to learn more about this concept, I encourage you to go through the following links:

- https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression
- http://data.library.virginia.edu/fitting-and-interpreting-a-proportional-odds-model

Thanks!

Machine Learning | Logistic Regression | Statistics | Econometrics

Predictive Analytics

Follow

Written by Akanksha Rawat

118 Followers · Writer for Towards Data Science

Graduate Research Assistant, Carnegie Mellon University

More from Akanksha Rawat and Towards Data Science

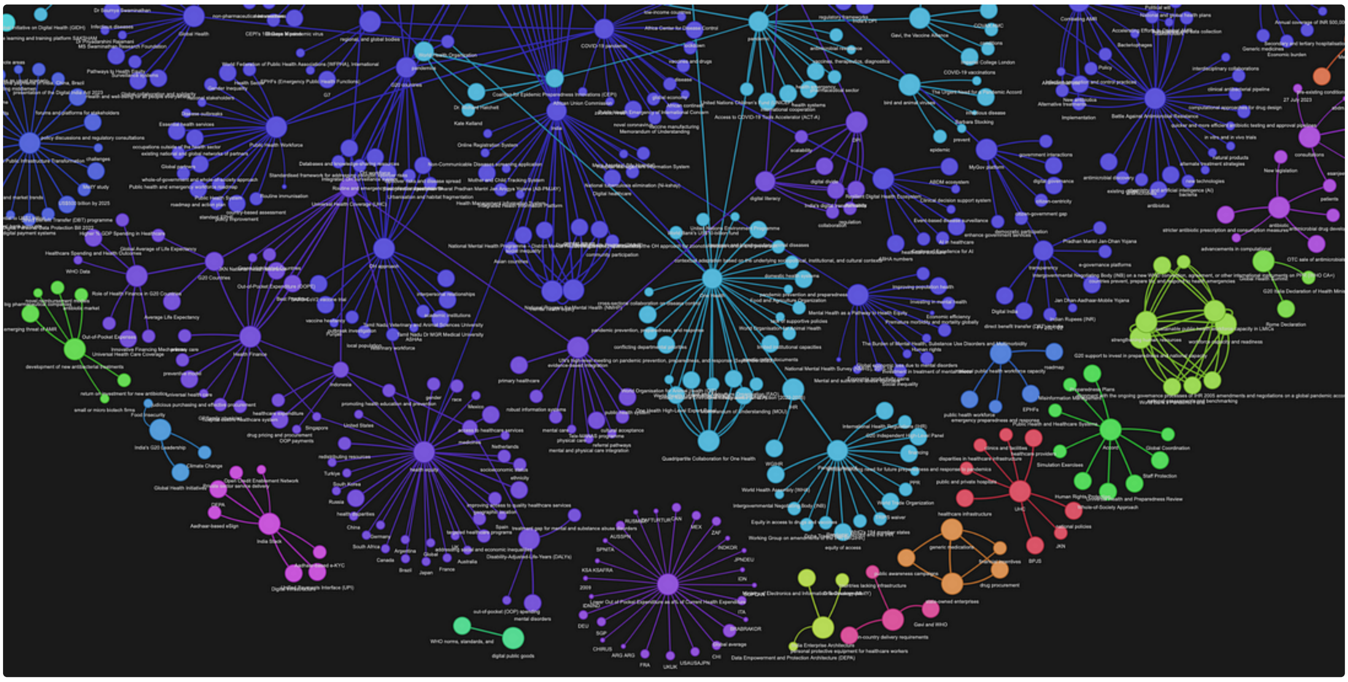
Akanksha Rawat in Towards Data Science

Binary Logistic Regression

An overview and implementation in R

6 min read · Oct 31, 2017

292 2



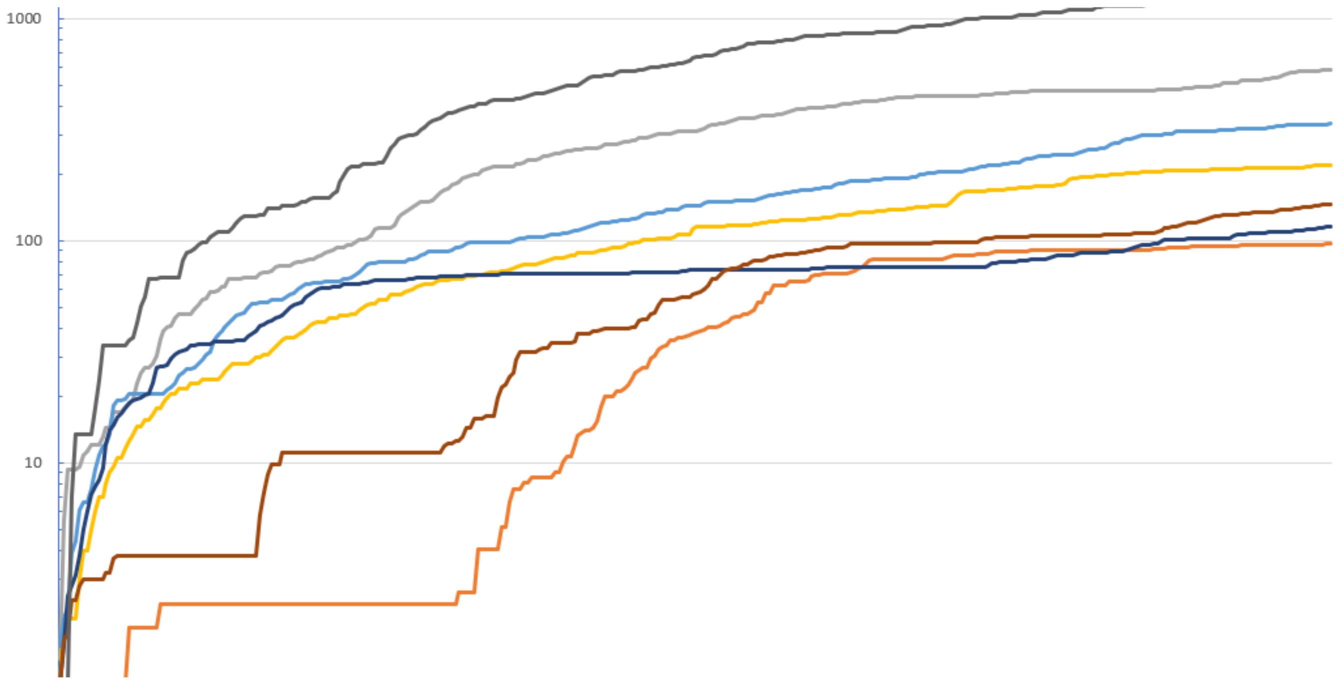
Rahul Nayak in Towards Data Science

How to Convert Any Text Into a Graph of Concepts

A method to convert any text corpus into a Knowledge Graph using Mistral 7B.

12 min read · Nov 10

4.2K 41



Pau Blasco i Roca in Towards Data Science

My Life Stats: I Tracked My Habits for a Year, and This Is What I Learned

I measured the time I spent on my daily activities (studying, doing sports, socializing, sleeping…) for 332 days in a row.

12 min read · Nov 20

4.4K 83

Anthony Alcaraz in Towards Data Science

Embeddings + Knowledge Graphs: The Ultimate Tools for RAG Systems

The advent of large language models (LLMs) , trained on vast amounts of text data, has been one of the most significant breakthroughs in…

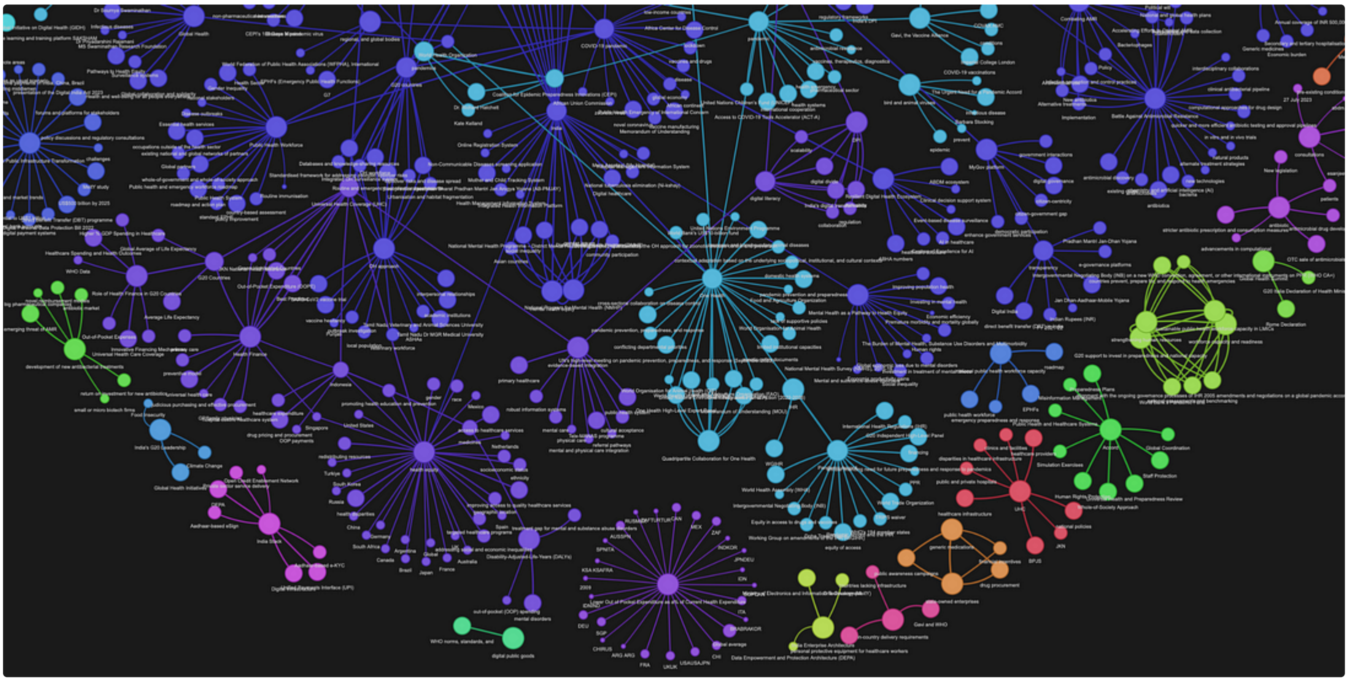
· 10 min read · Nov 14

1K 10

See all from Akanksha Rawat

See all from Towards Data Science

Recommended from Medium



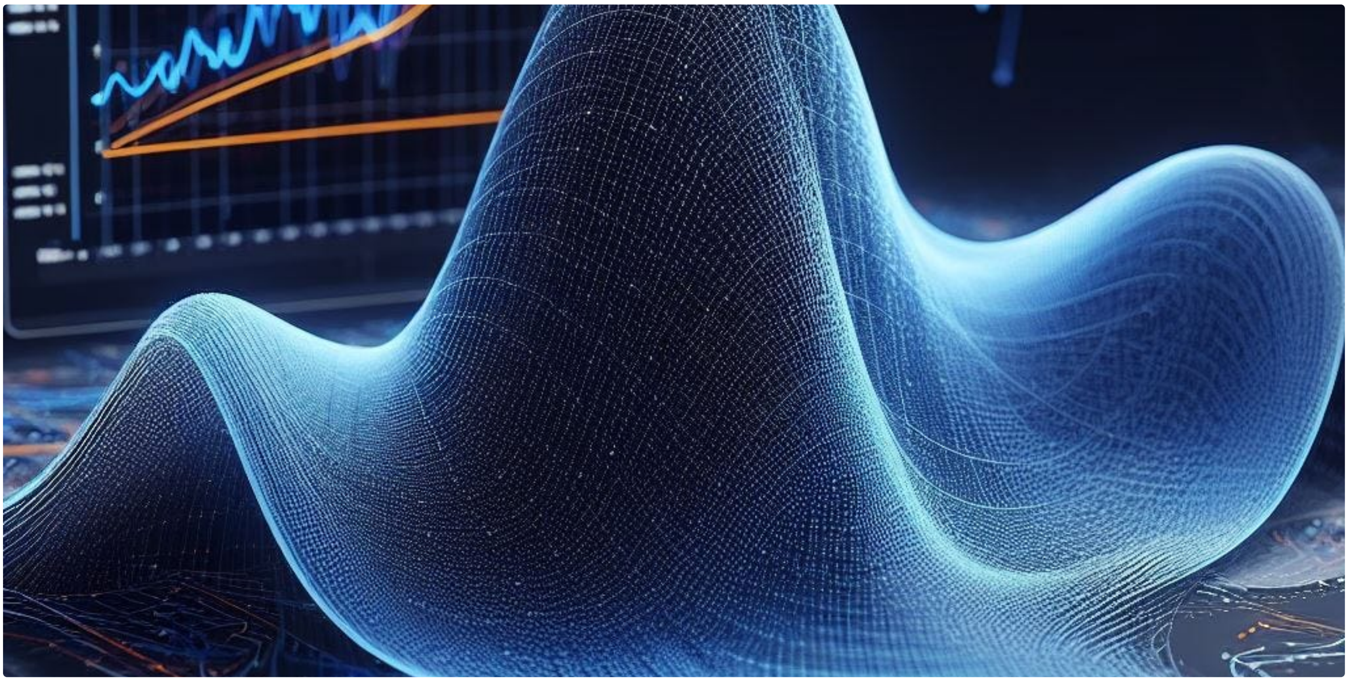
Rahul Nayak in Towards Data Science

How to Convert Any Text Into a Graph of Concepts

A method to convert any text corpus into a Knowledge Graph using Mistral 7B.

12 min read · Nov 10

4.2K 41



Bhanuprakash D

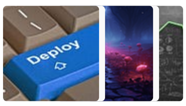
Logit function — A primer

Introduction

4 min read · Nov 25

11

Lists

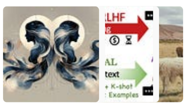


Predictive Modeling w/ Python

20 stories · 682 saves

Practical Guides to Machine Learning

10 stories · 776 saves

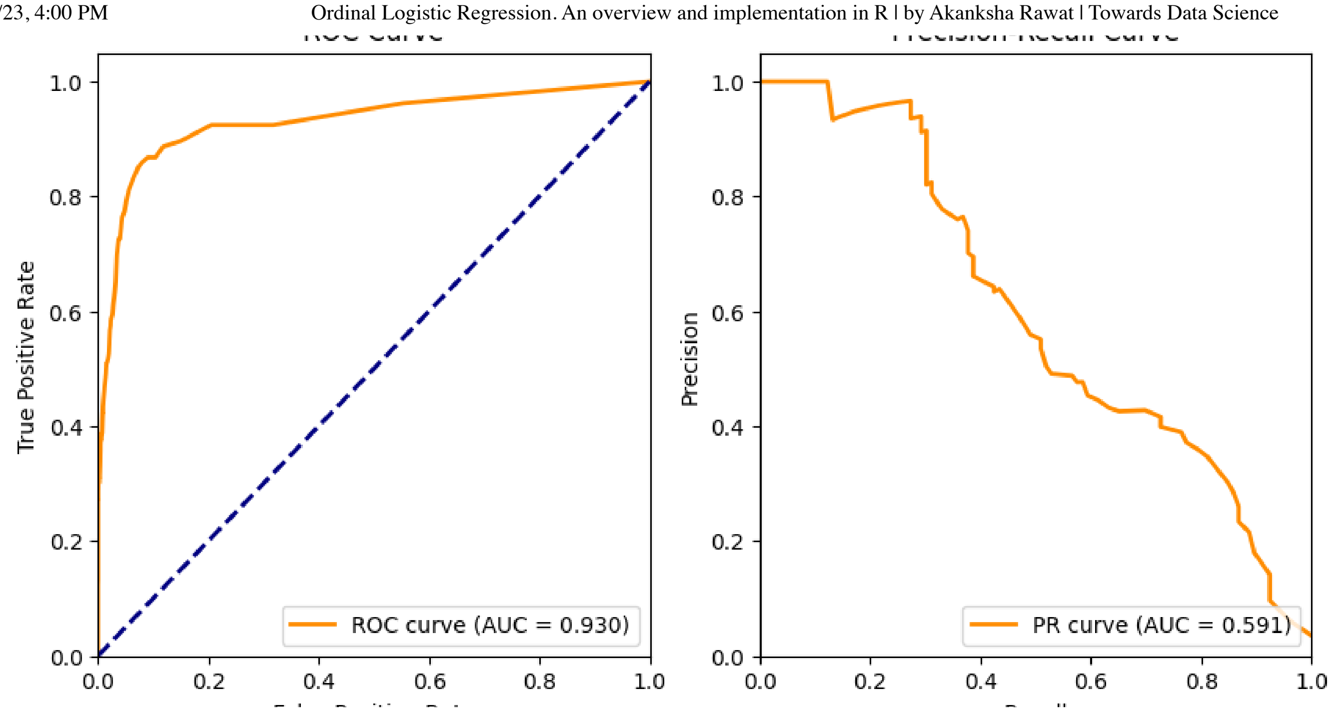


Natural Language Processing

970 stories · 463 saves

The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 234 saves



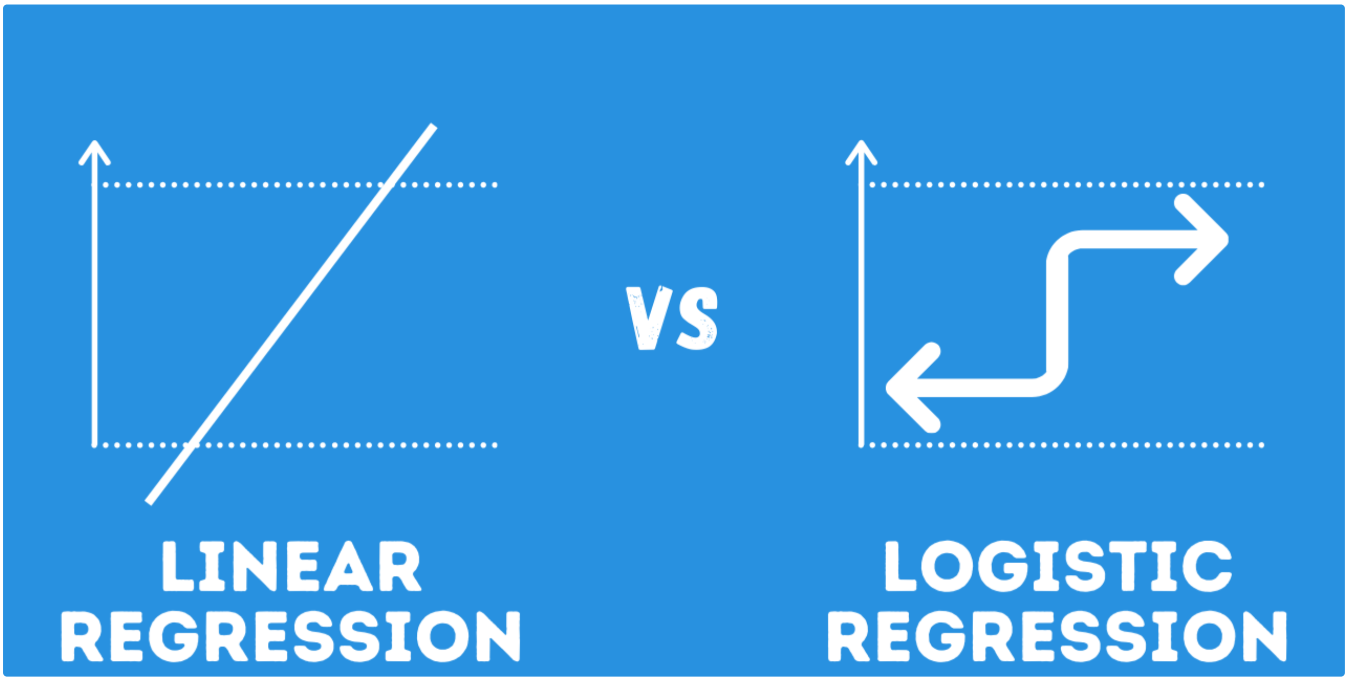
Juan Esteban de la Calle

How and Why I Switched from the ROC Curve to the Precision-Recall Curve to Analyze My Imbalanced...

Machine learning is transforming how we tackle problems and make decisions in an array of sectors, from healthcare to finance. Binary...

5 min read · Jul 10

55



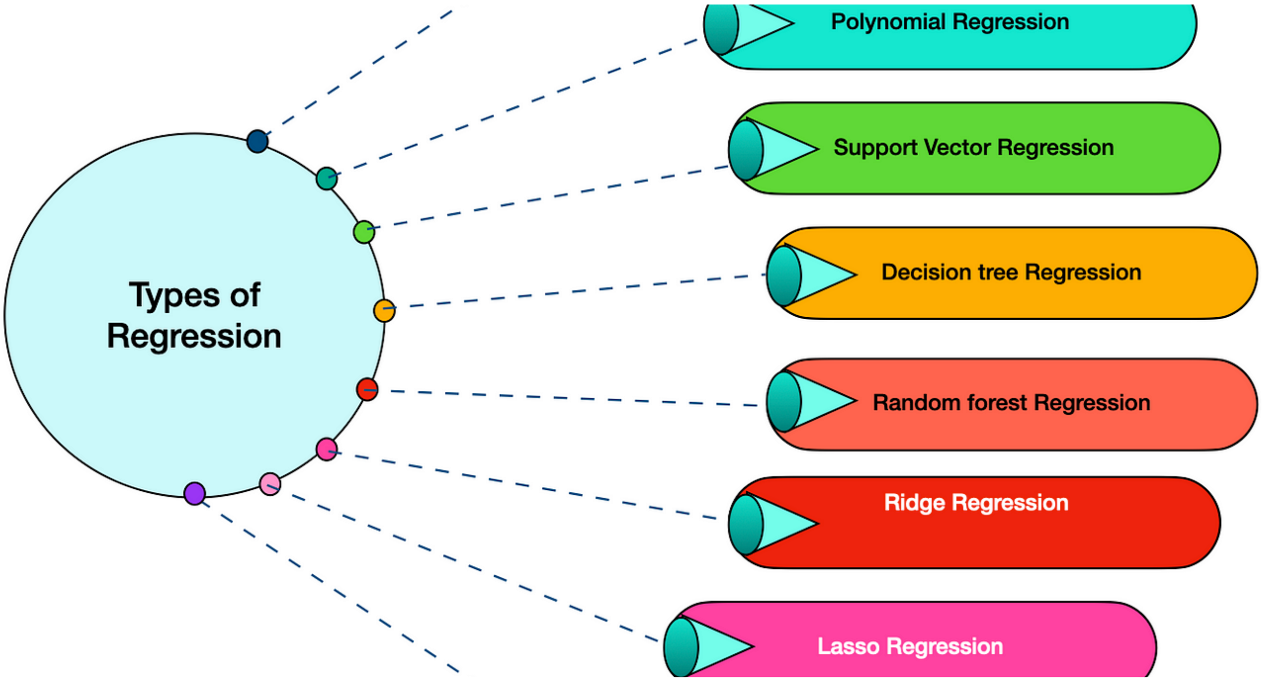
Praveen Raj

Why Is Logistic Regression Called “Logistic Regression” And not a Logistic Classification?

Since the target column is categorical, it uses linear regression to create an odd function that is wrapped with a log function to use...

2 min read · Aug 2

2



Arun Kumar Pandey (Ph.D., Astrophysics)

Regression algorithms

Regression algorithms are a type of machine learning algorithm used to predict numerical values based on input data. Regression algorithms...

5 min read · Jun 22

75

theDataGuy

Iris classification using Logistic Regression

Description: This is my first machine learning project :). This is a basic classification problem and the dataset used here is iris...

7 min read · Sep 28

1

See more recommendations