



Logistic Regression



What is Machine Learning?

- Machine learning is an area of artificial intelligence (AI) and computer science that focuses on using data and algorithms to mimic the way human beings learn, with the goal of steadily increasing accuracy.
- It is one of the key component of the rapidly expanding field of data science.
- It can solve problems at a speed and scale that the human mind cannot duplicate.



Usage of Machine Learning

- Data Management and Security
- Health Care
- Search Engines
- Finance
- Retail
- Image and Voice recognition
- Social Media Platforms
- Video Surveillance
- GPS Service
- Fraud Detection



Types Of Machine Learning

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement



What is Logistic Regression?

- Logistic regression is a statistical analytic method that uses past observations of a data set to predict a binary result, such as yes or no, true or false, etc.
- In order to model a binary output variable, logistic regression utilizes a logistic function.
- It is a classification model rather than a regression model
- The major distinction between linear regression and logistic regression is that the range of logistic regression is limited to values between 0 and 1.
- Can be used to analyze which variables influence the probability of a given outcome.

A short experiment



Video 1: Two groups passing basket balls [6]



Background of the experiment

- This is one of the parts of a psychology research done by Daniel Simons and Christopher Chabris at Harvard in 1999.
- 108 Subjects.
- They discovered that the more difficult the task, the more probable it is that people would miss the gorilla.
- 50% of the subjects spotted the gorilla.



Using Logistic Regression

Problem: We want to predict whether a person with a specific set of characteristics(independent variable) will detect the gorilla or not

Independent Variable: Concentration Span (scale 1-100)

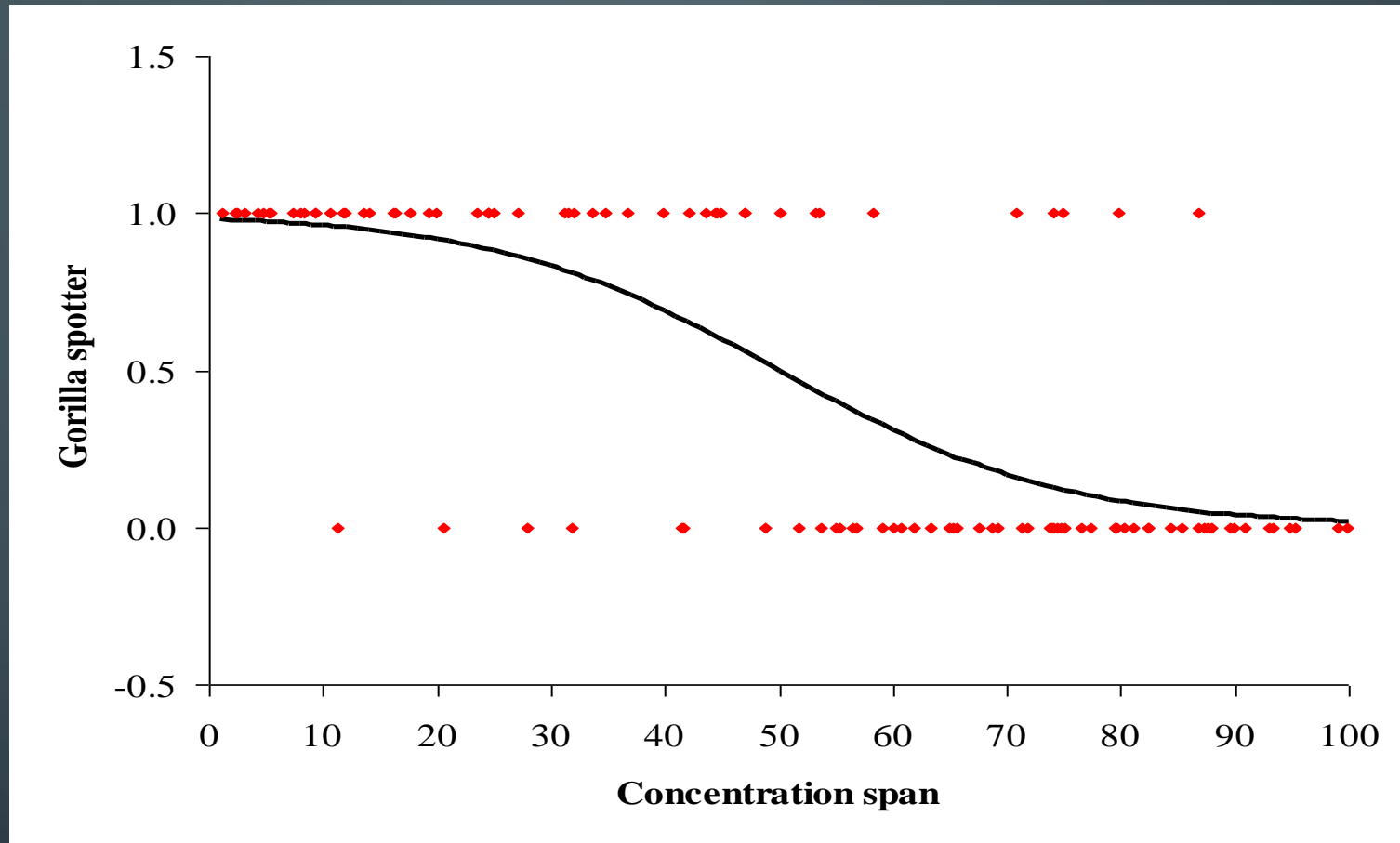


Figure 1: Graph representing the link between the probability of spotting the gorilla and CS

Simple Logistic Model Formula

$$\log\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta X$$

Logistic Regression Results Report

Variable	Coefficient value	Standard error	<i>p</i> -value
Concentration	-0.07	0.01	0.00
Intercept	3.69	0.72	0.00



Advantages of Logistic Regression

- Easy to interpret, implement and analyze.
- Allows to simply update models to incorporate new data.
- Provides accurately well-calibrated probability as well as classification results.
- Less prone to over-fitting in a low dimensional dataset.
- Using a softmax classifier, this algorithm can easily be augmented to multi-class classification.



Disadvantages of Logistic Regression

- Logistic regression cannot tackle nonlinear problems since it has a linear decision surface.
- It can only predict discrete functions.
- Logistic Regression necessitates either average or no multicollinearity between independent variables.
- Complex relationships are hard to determine when using logistic regression.



Applications of Logistic Regression

- Medical Sector
- Credit Score
- Gaming
- Spam Detection
- Bookings
- Social experiments and statistics

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a dark blue background, resembling a circuit board or a neural network diagram.

Conclusion

References

1. Peng, Chao-Ying Joanne, Kuk Lida Lee, and Gary M. Ingersoll. "An introduction to logistic regression analysis and reporting." *The journal of educational research* 96.1 (2002): 3-14.
2. Simons DJ, Chabris CF. Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*. 1999;28(9):1059-1074. doi:10.1068/p281059
3. Edgar, Thomas, and David Manz. *Research methods for cyber security*. Syngress, 2017.
4. Wright, Raymond E. "Logistic regression." (1995).
5. <https://www.expert.ai/blog/machine-learning-definition/>
6. <https://www.youtube.com/watch?v=Ahg6qcgoy4>