Machine Learning

Week 3 - Logistic Regression

About the course

Announcements

- Assignment for this week is already in the Notebooks folder
- Notebooks can be done as group (2-3 ppl)
- Two walkthrough videos for terminal (Mac and Windows) and Conda Navigator (Windows)
- This week's assigned paper is kaggle_competition_automated_essay_grading_…
- You can download this week's Notebook folder if you haven't and start Jupyter Notebook
- If you can, camera on is appreciated

Previously on ML..

- **What are the 3 main Machine Learning categories?**
- **What is the formula for a typical simple linear regression?**
- **Which expression of the Cost function do we prefer and why?**

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \qquad J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

- **How do we evaluate the performance of a simple linear regression?**
- **What is the difference between simple, multiple and polynomial regression?**
- **How can you improve a linear regression model?**

Logistic Regression

Definition

Logistic regression is a statistical **model** that in its basic form uses a **logistic** function to **model** a binary dependent variable.

Understanding the odds and log of odds

https://www.youtube.com/watch?v=ARfXDSkQf1Y&ab_channel=StatQuestwithJoshStarmer

The Derivative of Cost Function for Logistic Regression

https://medium.com/analytics-vidhya/derivative-of-log-loss-function-for-logistic-regression-9b832f025c2d#:~:text=Since%20the%20hypothesis%20function%20for,function%20follows%20a%20certain%20pattern.

https://rpubs.com/dnuttle/ml-logistic-cost-func_derivative

https://web.stanford.edu/class/archive/cs/cs109/cs109.1178/lectureHandouts/220-logistic-regression.pdf

Maximum Likelihood

https://www.youtube.com/watch?v=BfKanl1aSG0&ab_channel=StatQuestwithJoshStarmer

Conclusion