

Perceptron Classification: Aspiring Minds' Employability Outcomes

Raj Choudhary*

*School of Computer Science and Engineering, VIT Chennai, Tamil Nadu, India 600127

Email: raj.choudhary2016@vitstudent.ac.in *

Abstract: Perceptron is used to classify the salary slabs of the Aspiring Minds' Employability Outcomes 2015 dataset. Perceptron is loosely based on the biology of neurons, which fire when the input crosses a certain threshold. However, a perceptron generates only linear hypotheses and fails to classify non-linear data. The AMEO dataset characterizes non-linearity and tremendous overlap and hence, we find that a perceptron fails in the classification tasks.

Index Terms—Employability, Classification, Perceptron, Salary, Non-linear

1. Introduction

In this paper, salary slab classification is attempted using the AMEO dataset which contains the employability outcomes of engineering graduates. The features include standardized test scores, academic performance metrics and other background details of the candidate. The target variable is the first salary of the candidate. Perceptron learning is used to classify whether the candidate falls in one of the four salary slabs. One-vs-all technique is used to generate a classifier for each category of salary slab.

2. Methodology

A single perceptron is used to classify the data. The perceptron has a bias and a weighted unit which accepts input. If the weighted sum crosses the threshold, the perceptron fires. Perceptron produce hypotheses in the linear space. And hence, fails to classify non-linear data single-handedly. However, multiple perceptrons can be arranged to form a multilayer neural network which then gains the ability to classify non-linear data effectively.

3. Dataset – AMEO 2015

For every engineer, AMEO [1] dataset provides anonymised bio data information along with their respective skill scores and employment outcome information. Specifically, the following information is available for every engineer:

- 1) Scores on Aspiring Minds' AMCAT - a standardized test of job skills. The test includes cognitive, domain and personality assessments.

- 2) Personal information like gender and date of birth.
- 3) Pre-university information like 10th and 12th grade marks, board of education and 12th grade graduation year.
- 4) University information like GPA, college major, college reputation proxy, graduation year and college location.
- 5) The following employment outcome information is available for every engineer: First job, annual salary, First job title, First job location, Date of joining and leaving of first job.

AMEO 2015 has gained traction since its public release. Aspiring Minds annually publishes the National Employability Report, a data-driven commentary on graduates and their employability. A recent NER was based on an extension of this dataset.

4. Algorithm

Perceptron equation:

$$y = \sum_{j=1}^M x_j w_j + w_o \quad (1)$$

computed and the weight w_j is updated:

$$w_j^{(k+1)} = w_j^{(k)} + \eta(y_i - \hat{y}_i^{(k)})x_{ij} \quad (2)$$

The perceptron algorithm starts with an initial guess $w_1 = 0$ for the halfspace, and does the following on receiving the example x_i :

- 1) Predict $\text{sign}(w_i \cdot x)$ as the label for example x_i .
- 2) If incorrect, update $w_{i+1} = w_i + l(x_i)x_i$ else $w_{i+1} = w_i$.

5. Conclusion

Here, we confirm that a single perceptron fails to classify data that is not linearly separable. The AMEO dataset includes overlapping points and large number of features which are not suitable for a simple algorithm like perceptron.

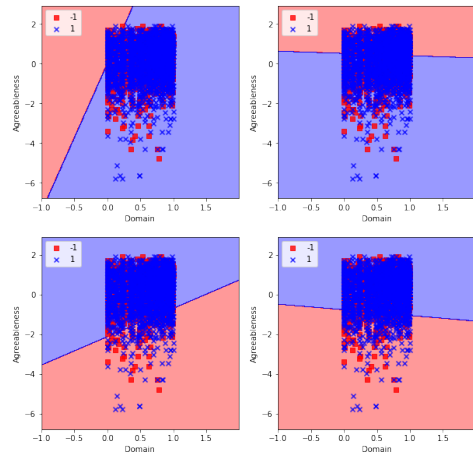


Figure 1. Implementation of perceptron

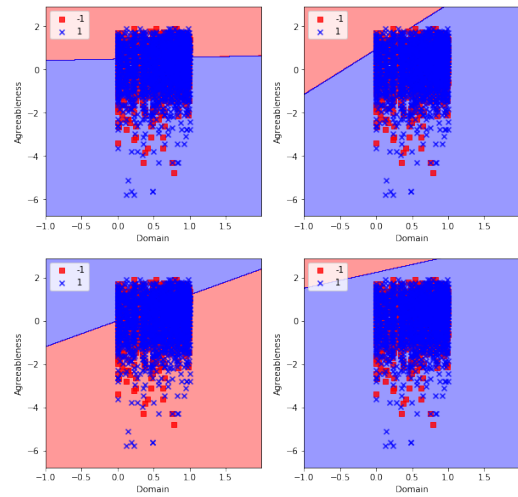


Figure 2. scikit-learn perceptron

6. References

- [1]. V. Aggarwal, S. Srikant, and H. Nisar, "Ameo 2015: A dataset comprising amcat test scores, biodata details and employment outcomes of job seekers," in *AMEO 2015: A Dataset Comprising AMCAT TestScores, Biodata Details and Employment Outcomes of Job Seekers*. ACM, 2016.