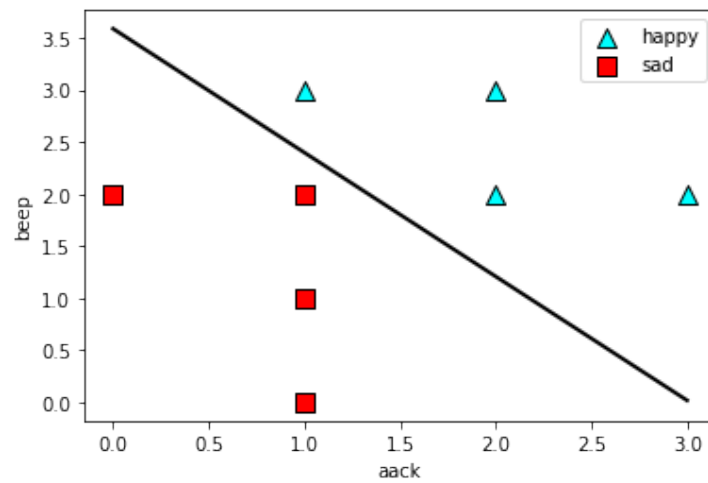


# COMP-2704: Supervised Machine Learning



## Assignment 3: The Perceptron

## Setup

Complete the following steps to setup this assignment:

- i. Open Jupyter Notebook in a web browser.
- ii. Within the “SupervisedML” folder, create a folder named “Assignment3” and navigate into this folder.
- iii. Open the link:  
[https://github.com/luisguiserrano/manning/tree/master/Chapter\\_5\\_Perceptron\\_Algorithm](https://github.com/luisguiserrano/manning/tree/master/Chapter_5_Perceptron_Algorithm)
- iv. Download the file:
  - a. *Coding\_perceptron\_algorithm.ipynb*
  - b. *utils.py*
- v. Download the file *SMLCode.exe* from the course website and run it. This will create a CSV file containing data.
- vi. Move all three files to the folder you created “~/SupervisedML/Assignment3”.
- vii. Open the *Coding\_perceptron\_algorithm.ipynb* notebook and run all code cells. Fix any errors that occur.

## Problem

For the following questions you will import data from the CSV file created in the Setup section above. This data represents the condition of an electrical mini-substation based on readings from voltage, current and temperature sensor readings. A condition of ‘o’ represents a properly functioning device, while a condition of ‘1’ represents failure. The model is to be used to alert maintenance staff of when a device has failed so that they can repair the mini-substation as soon as possible.

Create a notebook with filename *SML\_a3.ipynb* within the folder “~/SupervisedML/Assignment3”. Add code and markdown cells to complete the following steps. You may copy relevant lines of code from *Coding\_perceptron\_algorithm.ipynb*.

- 1) [2 mark] Which is worse for this use case, a false positive or a false negative? What value of  $\beta$  would be suitable for an  $F_\beta$  score?
- 2) [1 mark] Load the CSV file into an SFrame named *data*. Print the SFrame. Split the data into training/validation/testing sets using 80%/10%/10% respectively.
- 3) [2 marks] Is feature rescaling turned on by default for the function *turicreate.logistic\_classifier.create*? What scale (original or rescaled) are the coefficients given in?
- 4) [3 marks] Create perceptrons using Turicreate to classify *data* with ‘Condition’ as the target. Experiment with different values of hyperparameters to develop two different models.

- 5) [4 marks] For each model:
  - i) display the training and validation accuracies;
  - ii) display the confusion matrix on the validation set;
  - iii) calculate and display recall, precision, and  $F_\beta$  score (using the value of  $\beta$  you chose above) on the validation set.
- 6) [1 mark] Select which of your two models is the best (or declare a tie) and justify your choice by commenting on metrics and the confusion matrix.
- 7) [2 marks] Using the test set and your choice of best model:
  - i) calculate and display the accuracy;
  - ii) display the confusion matrix;
  - iii) calculate and display recall, precision, and  $F_\beta$  score.
- 8) [2 marks] With the exception of import and print statements, add a comment before each line of code in *SML\_a3.ipynb* to explain what it does.

## Contributions

Using markdown at the end of the notebook, list the contribution of each student to the assignment, referencing specific question numbers and other tasks such as formatting and submitting. Ideally, both students will contribute to all questions.

## Submission

Upload your notebook to the Assignment 3 dropbox on the course website. Late submissions will lose 10%.

**Total marks = 17**