02/04/25 LAB-4

Consider binary classification problem where we want to predict whether a student will pass or fail based on their study hrs. The logistic regression model has been trained & the learned parameters are $a_0 = -5$, $a_1 = 0.8$.

a) Write the logistic Regression equation for this problem

$$p(y=1|x) = \frac{1}{1+e^{-(-5+0.8x)}}$$

b) Calculate the probability that a student studies for 7 hrs will pass

$$p(\text{pass}) = \frac{1}{1+e^{-0.6}} = 0.6479$$

c) Determine the predicted class for this student based on threshold 0.5. -> if $p(\text{pass}) \geq 0.5$ -> student will pass otherwise he will fail

d) Consider $z = [2, 1, 0]$ for three class. Apply softmax function to find the probability values of three classes

$$\text{softmax}(z_0) = \frac{e^{z_0}}{\sum_{i=0}^{k} e^{z_i}}$$

$$p(1) = \frac{e^2}{e^2+e^1+e^0} = 0.665$$

$$p(2) = \frac{e^1}{e^2+e^1+e^0} = 0.245$$

$$p(3) = \frac{e^0}{e^2+e^1+e^0} = 0.09$$

Q. Dataset file : "HR_comma_sep.csv"

i) Which vars did you identify as having a direct & clear impact on emp retention & why?

-> satisfaction level, time spent in company, no of proj, salary

These variables are choosen based on the trends in data visualization

ii) What was the accuracy of your logistic regression model? Do you think this is a good accuracy? Why or why not?

-> The accuracy of logistic regression was 78%. This accuracy is fairly good. It suggests that model captures most of properties affecting employee retention.

2) For zoo dataset

i) Did you perform any data preprocessing steps? If yes, what were they and why were they necessary?

-> Yes, -> Dropped 'animal name' column, checked for missing values, converted categorical variable if needed

ii) Were there any missing or inconsistent values in the dataset? How did you handle them?

-> No missing values were found in the dataset. If they were present/inconsistent, we could use mean/mode methods to correct them.

iii) What does the confusion matrix tell you about the performance of your model?

-> It shows how well the model predicted different class types. Animals.

iv) Most frequently misclassified class: some class types are misclassified. This happens due to similarity in features b/w different animal classes.