

Assignment 1: Classification Losses

40 Points

Deliverables: Answers to background questions; edited Excel spreadsheet; Python code in an .ipynb file.

Data:

binaryPrediction.csv: Table of class labels and predictions for a binary classification problem. These data are synthetic and do not represent a real classification problem.

Variable	Description
sampleID	Unique ID for each sample
refLabel	Reference class label
refCode	Reference numeric code
predLogit	Predicted class logit

multiclassPrediction.csv: Table of class labels and predictions for a 4-class problem. These data are synthetic and do not represent a real classification problem.

Variable	Description
sampleID	Unique ID for each sample
refLabel	Reference class label
refCode	Reference numeric code
logitA	Predicted logit for Class A
logitB	Predicted logit for Class B
logitC	Predicted logit for Class C
logitD	Predicted logit for Class D
predCode	Predicted numeric class code
predLabel	Predicted class label

Background Questions

B1: Write the formula for binary cross entropy loss. Explain each term in the equation. (4 Points)

B2: Write the formula for cross entropy loss. Explain each term in the equation. (4 Points)

B3: Explain the difference between a logit and a class probability. (4 Points)

B4: A binary classification can be framed as a two-class problem (e.g., “presence” and “absence”) or a one-class problem (e.g., “presence”). Explain how this decision impacts (1) the design of the neural network architecture (specifically, the final output) and (2) the appropriate loss metric to apply. (4 Points)

B5: As implemented in PyTorch, explain the difference between negative log likelihood loss and cross entropy loss. (4 Points)

Tasks (Excel):

Perform the following calculations in Excel.

T1: For the binary classification results, convert the raw logits to estimated class probabilities relative to the positive case. (4 Points)

T2: Use the estimated positive class probability for each sample to calculate the binary cross entropy loss for the classification. (4 Points)

T3: For the multiclass classification, convert the class logits to estimated probabilities for each class. (4 Points)

T4: Use the class probabilities and class labels to calculate the cross entropy loss for the classification. Note that you will need to generate one-hot encodings for the correct class labels. (4 Points)

Tasks (PyTorch):

Perform the following calculations using Python/PyTorch.

T5: Convert the data into PyTorch tensors and edit as necessary. Calculate the binary cross entropy loss for the binary classification and the cross entropy loss for the multiclass classification using the loss functions implemented in PyTorch to check your work in Excel. (4 Points)