Sample Questions for the Final Exam Performance Estimation

Question 1:

A binary classification model produces the following results on a test set:

- True Positives (TP): 40
- False Positives (FP): 10
- False Negatives (FN): 17

Calculate the model’s F1-score.

A. 0.80

B. 0.75

C. 0.71

D. 0.67

Soluiton

- Precision=TP/(FP+TP) =40/(10+40)=0.8
- Recall=TP/(FN+TP)=40/(17+40)≈0.7018
- F1=2×Precision×Recall/(Precision + Recall) ≈**0.7476**

Question 2:

A deep learning model is trained to classify images into three categories: Dog, Wolf, and Cat. On the test set, the following results were observed:

Class	F1-score	True Positives (TP)	False Positives (FP)
Dog	0.75	40	20
Wolf	0.60	30	30
Cat	0.80	50	10

Calculate the model's Balanced Accuracy.

A. 0.86
B. 0.79
C. 0.76
D. 0.72

Solution

- Step 1: What is the **Balanced Accuracy** (*Macro Recall*)?
- Step 2: Recall = TP/(TP+FN). We don't have FN directly, but we can derive recall from F1 and precision, and we can calculate precision from TP and FP:
- Step 3: Precision = TP/(TP+FP)
  - Precision_dog = 40/(40+20) = 0.667; Precision_wolf = 30/(30+30) = 0.5 ;
  - Precision_cat = 50/(50+10) = 0.833
- Step 4: Use **F1 and Precision** to find **Recall**
  - **F1 = 2*** Precision* Recall /(Precision + Recall) ➔ Recall = F1* Precision/(2 Precision-F1)
  - Recall_dog ≈0.857; Recall_wolf =0.750; Recall_cat = ≈0.769
- Step 5: Balanced Accuracy= (Recall_dog + Recall_wolf + Recall_cat) /3

  = (0.857+0.750+0.769) /3 ≈**0.792**

Question 3:

In **multiclass classification**, where each sample has a **single label**, when using **micro-averaging**, what is the correct order of the following metrics?

A. F1-score = Accuracy, Recall ≠ Precision
B. Accuracy ≥ F1-score ≥ Recall ≥ Precision
C. F1-score = Accuracy = Recall = Precision
D. F1-score ≠ Accuracy, Recall = Precision

Solution

In multiclass classification with micro-averaging, precision, recall, and F1-score are computed by aggregating all true positives, false positives, and false negatives across all classes.

Since the total number of false positives equals the total number of false negatives, micro-precision equals micro-recall, and their harmonic mean (the micro F1-score) equals both.

Furthermore, since accuracy is defined as the number of correct predictions over the total number of predictions, and micro-averaging also computes metrics over all predictions, accuracy equals micro-precision, micro-recall, and micro-F1.

In all these metrics, the numerator is the sum of T_i (i.e., the diagonal elements of the confusion matrix), and the denominator is the total number of samples *N* (i.e., the sum of all entries in the confusion matrix). Therefore: F1-score = Accuracy = Recall = Precision.