SFU

University of LaTeX

Department of Computer Science

Homework 1

CS 101: Introduction to Computer Science

Professor Jane Doe

John Smith - 123456789

Due Date September 30, 2025

Fall 2025

Question 1

Preliminaries:

We consider a binary classification task with two classes: positive (*pos*) and negative (*neg*). The following quantities are defined:

- N_{pos}: Number of positive instances,
- N_{neg}: Number of negative instances,
- $r = \frac{N_{\text{neg}}}{N_{\text{pos}}}$: Class imbalance ratio (negatives to positives).

A classification outcome can be summarized via the confusion matrix:

		Predicted	
		Positive	**Negative**
Actual	**Positive**	TP	FN
	Negative	FP	TN

We define the following evaluation metrics:

- **Recall (True Positive Rate)**

$$\text{Recall} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{N_{\text{pos}}}.$$

- **Precision**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

- **F1 Score**

$$F_1 = 2\,\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

- **False Positive Rate**

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{N_{\text{neg}}}.$$

For different thresholds on the classifier score, we define the following plots:

- **PR Curve:** Precision vs. Recall.
- **ROC Curve:** TPR vs. FPR.

Solution

Let us fix the TPR $= \alpha$ and FPR $= \beta$ as they are the properties of the classifier. These two metrics are *normalized* by the class sizes N_{pos} and N_{neg}, respectively:

$$\alpha = \frac{\text{TP}}{N_{\text{pos}}}, \qquad \beta = \frac{\text{FP}}{N_{\text{neg}}}.$$

Because they measure *rates* of correct or incorrect classification within each class, α and β do not explicitly depend on the ratio r. Consequently, for a chosen threshold (i.e. a chosen point on the ROC curve), α and β can remain rather constant even as we vary the class imbalance ratio $r = \frac{N_{\text{neg}}}{N_{\text{pos}}}$.

Parametrizing Confusion Matrix Entries:

Assume $N_{\text{neg}} = r\, N_{\text{pos}}$. Then the confusion matrix would be:

		Predicted	
		Positive	**Negative**
Actual	**Positive**	$\alpha\, N_{\text{pos}}$	$(1-\alpha)\, N_{\text{pos}}$
	Negative	$\beta\, r\, N_{\text{pos}}$	$(1-\beta)\, r\, N_{\text{pos}}$

Using this, we derive the following for Precision, Recall, and F1 Score in terms of α, β, and r:

Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\alpha\, N_{\text{pos}}}{\alpha\, N_{\text{pos}} + \beta\,(r\, N_{\text{pos}})} = \frac{\alpha}{\alpha + \beta\, r}.$$

Recall (TPR):

$$\text{Recall} = \alpha \quad (\text{since } \alpha = \tfrac{\text{TP}}{N_{\text{pos}}}).$$

F1 Score:

$$F_1 = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2\,\frac{\left(\frac{\alpha}{\alpha+\beta\, r}\right)\alpha}{\left(\frac{\alpha}{\alpha+\beta\, r}\right) + \alpha} = \frac{2\,\alpha}{2\,\alpha + \beta\, r}.$$

Sensitivity of PR Metrics (Precision, F1) to Imbalance

- **Precision** $= \dfrac{\alpha}{\alpha + \beta r}$: As r grows large (i.e. many more negatives than positives), the term βr can dominate the denominator unless β is driven extremely close to zero. Thus, even if β is small, the large product $\beta\, r$ makes Precision plummet.
- **F1 Score**: Similarly, the F1 Score depends on r in the denominator. As r increases, the denominator grows, which reduces the F_1 score.

Hence, PR-based metrics explicitly reflect how the minority (positive) class is being detected relative to the potentially huge negative class. Small β can still yield large absolute FP, which greatly impacts Precision and thus reduces F_1.

ROC Curve: Insensitivity to r The ROC curve plots:

$$(\text{FPR}, \text{TPR}) \;=\; (\beta, \alpha).$$

Neither α nor β is multiplied by r. Thus for the *same* α and β, the ROC curve *does not change* as r varies. Even if r becomes very large, the FPR remains the same β. Consequently, one can have a deceptively high Area Under the ROC Curve (AUC) while still missing the minority class in absolute terms. This insensitivity arises because FPR is normalized by N_{neg}, so it remains "locally small" even if N_{neg} itself is huge.

To summarize:

- **PR curve and F1 score are sensitive to class imbalance.** The denominators of these metrics depend on the absolute number of false positives, which scales with N_{neg}. As $r = \frac{N_{\text{neg}}}{N_{\text{pos}}}$ increases, Precision and F1 score can drop significantly even if the FPR remains constant. Thus, these metrics are directly affected by class imbalance.
- **ROC curve is relatively insensitive to class imbalance.** The ROC curve plots $\text{TPR} = \frac{\text{TP}}{N_{\text{pos}}}$ against $\text{FPR} = \frac{\text{FP}}{N_{\text{neg}}}$. For fixed α and β, changing r does not alter the ROC point (β, α). Therefore, a model can appear to perform well on the ROC curve even under high class imbalance, potentially masking poor Precision and low F1 scores.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: 1512.03385 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1512.03385.