UNIVERSIDAD DE MÁLAGA

Aprendizaje Computacional

Lab 2A:
Classification Performance Analysis

AUTHORS:

JANECZKO TOMASZ

LUKÁŠ HOFMAN

November 7, 2024

1 Introduction to problem

Classification is a key task in machine learning: when we train such a model, the training set includes data points and their known class. It is important to see how these models perform in a practical scenario. This is the analysis of various classification methods (A, B, C, D and E). We will compare their performance with the help of well-established metrics like Precision (PR), Recall (RC), Specificity (SP), False Negative Rate (FNR), and False Positive Rate (FPR). We will use them to compare against existing methods using Accuracy (ACC), Spatial Accuracy (S), F-measure (Fm). By analyzing these metrics, we would try to find the best suitable method for that particular classification.

2 Dataset

Method	TP	FP	FN	TN
A	100	900	0	0
B	80	125	20	775
C	25	25	75	875
D	50	50	50	850
E	0	0	100	900

Fig. 1. Provided dataset from the pdf file

- **Number of Samples:** 1000
- **Number of Classes:** 2 (Positive and Negative)
- **Number of Samples:**
  - Positive Class: 100
  - Negative Class: 900

It is worth noting that for all methods we get the same number of positive and negative results, otherwise that wouldn't be the same dataset.
With only 10% of the dataset being actual positives, we can easily say that the dataset is imbalanced.

3 Metrics

As previously noted, we will be using in our project 8 different metrics.

3.1 Precision (PR)

- **Represents:** The accuracy of positive predictions.
- **Definition:** Precision is the ratio of true positives to the sum of true positives and false positives.
- **Formula:** PR = TP / (TP + FP)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, as they indicate fewer false positives among predicted positives.

3.2 Recall (Sensitivity or True Positive Rate) (RC)

- **Represents:** The ability to capture all actual positives.
- **Definition:** Recall is the ratio of true positives to the sum of true positives and false negatives.
- **Formula:** RC = TP / (TP + FN)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, as they indicate that the method is identifying most true positives.

3.3 Specificity (True Negative Rate) (SP)

- **Represents:** The ability to correctly identify negatives.
- **Definition:** Specificity is the ratio of true negatives to the sum of true negatives and false positives.
- **Formula:** SP = TN / (TN + FP)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, as they indicate that the method accurately identifies negatives with few false positives.

3.4 False Negative Rate (FNR)

- **Represents:** The proportion of actual positives missed by the method.
- **Definition:** FNR is the ratio of false negatives to the sum of true positives and false negatives.
- **Formula:** FNR = FN / (TP + FN)
- **Range:** [0, 1]
- **Interpretation:** Lower values are better, as they indicate fewer missed positives.

3.5 False Positive Rate (FPR)

- **Represents:** The proportion of actual negatives incorrectly identified as positives.
- **Definition:** FPR is the ratio of false positives to the sum of false positives and true negatives.
- **Formula:** FPR = FP / (FP + TN)
- **Range:** [0, 1]
- **Interpretation:** Lower values are better, as they indicate fewer false positives.

3.6 Accuracy (ACC)

- **Represents:** The overall correctness of predictions.
- **Definition:** Accuracy is the ratio of correctly predicted instances (true positives and true negatives) to the total number of instances.
- **Formula:** ACC = (TN + TP) / (TP + FN + FP + TN)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, indicating a higher overall correctness.

3.7 Spatial Accuracy (Jaccard Index) (S)

- **Represents:** The effectiveness of positive identification relative to all positive and false detections.
- **Definition:** Spatial Accuracy (S) measures the ratio of true positives to the sum of true positives, false negatives, and false positives, offering a focused accuracy measure for identifying positives.
- **Formula:** S = TP / (TP + FN + FP)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, as they indicate a stronger positive detection performance with minimal false positives and negatives.

3.8 F-measure (F1 Score) (Fm)

- **Represents:** The harmonic mean of precision and recall, balancing both metrics.
- **Definition:** F-measure combines precision and recall to give a single measure of a method's effectiveness.
- **Formula:** Fm = (2 * PR * RC) / (PR + RC)
- **Range:** [0, 1]
- **Interpretation:** Higher values are better, as they indicate a balanced approach between identifying positives and avoiding false positives.

4 Yielded performance

Method	Precision	Recall	Specificity	FNR	FPR	Accuracy	Spatial Accuracy	F-measure
A	0.10	**1.00**	0.00	**0.00**	1.00	0.10	0.10	0.18
B	0.39	0.80	0.86	0.20	0.14	0.86	**0.36**	**0.52**
C	**0.50**	0.25	0.97	0.75	0.03	**0.90**	0.20	0.33
D	**0.50**	0.50	0.94	0.50	0.06	**0.90**	0.33	0.50
E	0.00	0.00	**1.00**	1.00	**0.00**	**0.90**	0.00	0.00

Table 1. Performance Metrics for Different Methods

5 Analysis of each method

5.1 Method A

– Precision: **0.10** - Very low, meaning many predicted positives are false.
– Recall: **1.00** - Excellent at identifying actual positives, with no false negatives.
– Specificity: **0.00** - Poor, unable to identify true negatives.
– False Negative Rate: **0.00** - No missed positives.
– False Positive Rate: **1.00** - High, with many false positives.
– Accuracy: **0.10** - Very low due to the inability to identify true negatives.
– Spatial accuracy: **0.10** - Low, indicating poor overall positive identification accuracy.
– F-measure: **0.18** - Poor balance between precision and recall due to low precision.

5.2 Method B

– Precision: **0.39** - Moderate, with a decent balance between true positives and false positives.
– Recall: **0.80** - High, identifying most positives.
– Specificity: **0.86** - High, with effective negative identification.
– False Negative Rate: **0.20** - Low, missing only a small proportion of positives.
– False Positive Rate: **0.14** - Acceptable, with relatively few false positives.
– Accuracy: **0.86** - High, due to a balanced identification of positives and negatives.
– Spatial accuracy: **0.36** - Moderate, showing method B reasonably covers positives and still does not overguess.
– F-measure: **0.52** - Highest among all methods, reflecting a strong balance.

5.3 Method C

– Precision: **0.50** - High, with few false positives.
– Recall: **0.25** - Low, missing many positives.
– Specificity: **0.97** - Very high, with nearly all negatives identified correctly.
– False Negative Rate: **0.75** - Poor, missing most positives.
– False Positive Rate: **0.03** - Very low, with minimal false positives.
– Accuracy: **0.90** - High, mainly due to excellent specificity.
– Spatial accuracy: **0.20** - Lower due to too many false negatives.
– F-measure: **0.33** - Low, as method C has low recall.

5.4 Method D

– Precision: **0.50** - The highest precision, meaning fewer false positives.
– Recall: **0.50** - Moderate, catching half of the positives.
– Specificity: **0.94** - High, accurately identifying negatives.
– False Negative Rate: **0.50** - Average, missing half of positives.
– False Positive Rate: **0.06** - Low, with few false positives.
– Accuracy: **0.90** - High, due to good specificity and reasonable recall.
– Spatial accuracy: **0.33** - Moderate, due to the number of correctly guessed positives is the same as the number of false positives and the number of false negatives
– F-measure: **0.50** - Balanced, though slightly lower than Method B.

5.5   Method E

– Precision: **0.00** - Fails completely to make accurate positive predictions.
– Recall: **0.00** - Identifies no positives at all.
– Specificity: **1.00** - Perfect, identifying all negatives correctly.
– False Negative Rate: **1.00** - Misses all positives.
– False Positive Rate: **0.00** - No false positives, but no true positives either.
– Accuracy: **0.90** - High but misleading, only due to perfect negative predictions.
– Spatial accuracy: **0.00** - Zero as the method doesn't predict any positives
– F-measure: **0.00** - Fails to balance any meaningful prediction of positives.

6   Analysis of the performance

6.1   Precision (PR)

– **Best:** Methods C and D (0.5)
  These methods have the highest precision, meaning they have the highest percentage of predicted positives being truly positive.
– **Worst:** Method E (0.0)
  Method E does not predict any positives.

6.2   Recall (RC)

– **Best:** Method A (1.0)
  Method A has the highest recall, identifying all actual positives.
– **Worst:** Method E (0.0)
  Method E does not predict any positives.

6.3   Specificity (SP)

– **Best:** Method E (1.0)
  Method E has perfect specificity, identifying all actual negatives.
– **Worst:** Method A (0.0)
  Method A does not predict any negatives.

6.4   False Negative Rate (FNR)

– **Best:** Method A (0.0)
  Method A captures all actual positives, achieving a perfect false negative rate.
– **Worst:** Method E (1.0)
  Method E misses all actual positives, achieving the worst possible false negative rate.

6.5   False Positive Rate (FPR)

– **Best:** Method E (0.0)
  Method E captures all actual negatives, achieving a perfect false positive rate.
– **Worst:** Method A (1.0)
  Method A misses all actual negatives, achieving the worst possible false positive rate.

6.6   Accuracy (ACC)

– **Best:** Methods C, D, and E (0.9)
  These methods achieve the highest accuracy, mainly due to strong specificity.
– **Worst:** Method A (0.1)
  Method A has the lowest accuracy, as only 10% of the test data are actual positives.

6.7 Spatial Accuracy (S)

– **Best:** Method B (0.356)
  Method B has the highest spatial accuracy, effectively balancing the identification of positives and negatives.
– **Worst:** Method E (0.0)
  Method E has no spatial accuracy, failing to predict any positives.

6.8 F-measure (Fm)

– **Best:** Method B (0.525)
  Method B provides the best balance between precision and recall, making it effective at identifying positives with fewer false positives.
– **Worst:** Method E (0.0)
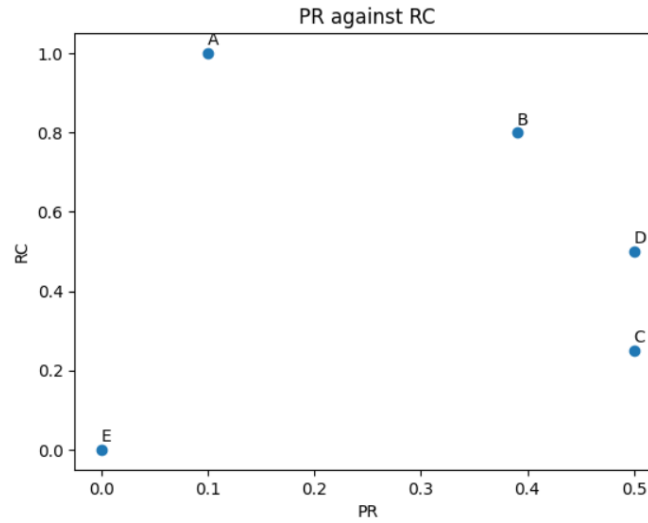 With no true positives predicted, Method E fails to achieve any meaningful balance.

7 Graphs

7.1 FN against FP
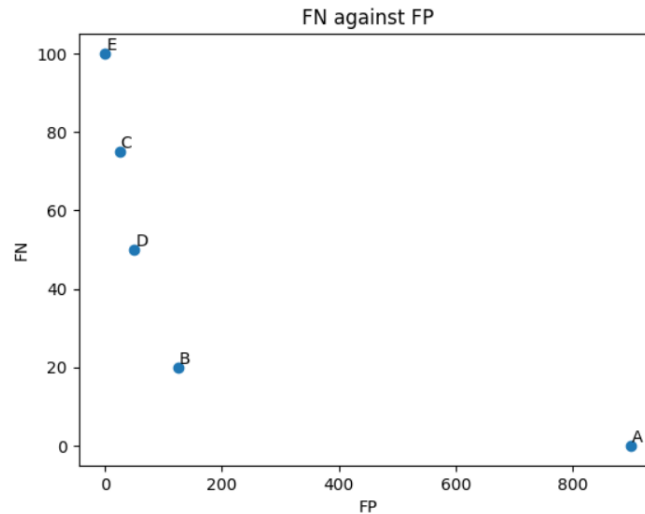
Fig. 2. Figure FN against FP

– **Best Method (B):**
  B is the best as it minimizes both FP and FN.
– **Worst Method (A, E):**
  A has high FP, incorrectly identifying many negative cases as positive. E has high FN, missing many actual positive cases.
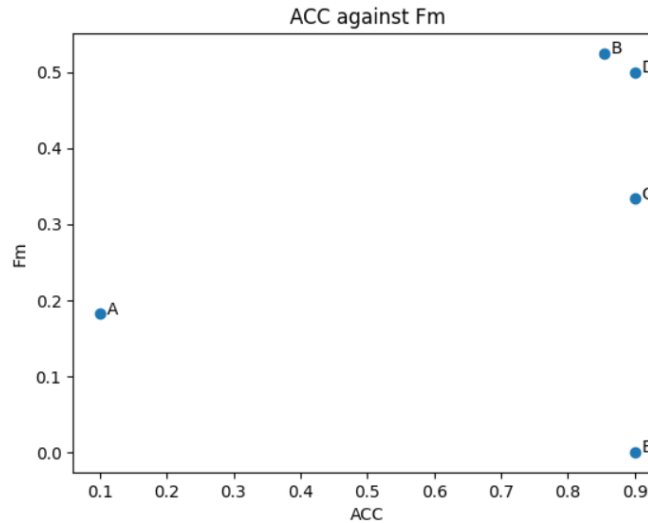
7.2 PR against RC

Fig. 3. Figure PR against RC

– **Best Method (B, D):**
  Both B and D have a good balance between PR and RC, making them reliable choices.
– **Worst Method (E):**
  E has a poor balance between PR and RC, resulting in completely inaccurate predictions.

7.3 ACC against Fm

Fig. 4. Figure ACC against Fm

– **Best Method (B, D):**
  B and D have a high values in both ACC and Fm, making them the best choices.
– **Worst Method (A, E):**
  A and E have either low ACC or low Fm, making them the least desirable options.

8   Conclusion

Method B stands out as the overall best performer, effectively balancing various metrics. It boasts the highest F-measure (0.52), indicating a strong compromise between precision (0.39) and recall (0.8). Additionally, it achieves a respectable accuracy of 0.85. This balanced performance makes Method B a solid choice, capable of accurately identifying positives while minimizing false positives.

However, the optimal method ultimately hinges on specific objectives. If absolute certainty in positive predictions is paramount, even at the cost of missing some true positives, Method C might be a more suitable option.