ROC Tool: Compute ROC curves and iso-PM ROC Curves.

Lavazza Luigi, Morasca Sandro, and Rotoloni Gabriele

Version: 20250809

ROC Tool is a tool that allows the user to visualize ROC curves and easily perform more advanced analysis on them. Specifically, the tool allows users to plot multiple ROC curves and their Regions of Interest, compute AUC and RRA, and plot multiple iso-PM curves.

This documentation file can be accessed by pressing the "Help" button inside the application.

For more information on how Regions of Interest, RRA, and iso-PM curves work, refer to the following paper: [XXXX]. If you use this tool for your research, please reference that paper.

Content

1 Installation and Execution

1.1 Repository structure

ROC Tool can be downloaded from GitHub (`kaal`). The repository contains the following files:

- **iRRA.zip** This is the R package needed to compute Regions of Interest (RoI) and RRA.The package was developed for the following paper **??**;

- **ROC˙Tool.R** Is the R file containing the Shiny application.

- **utils** Contains all other R files needed to perform all computations needed for iso-PM curves.

- **doc** Contains this PDF file.

1.2 Before running the tool

Before running the tool, it is necessary to install all needed packages. All packages except iRRA are automatically installed (if not already installed) when the tool is executed.

iRRA has to be manually installed. In RStudio, the user needs to go to `Tools -> Install Packages...`. In the dialog window, select Package Archive File in the Install from: menu, and then select the `iRRA.zip` file. It is possible to verify if the installation went well by executing `library(iRRA)`.

If the RStudio is not able to install iRRA this way, it is also possible to extract `iRRA.zip` and then move the extracted folder manually in the R folder that contains all the installed libraries (The user can see where the folder is with the `.libPaths()` function).

1.3 Execute the tool

To execute the tool the user needs to open the `ROC_Tool.R` file and press the "Run App" button on the top-right corner of the source code pane.

2 Performance Metrics Supported by the Tool

Given a confusion matrix like in Table 2, allows for the computation of multiple performance metrics (abbreviated in this document as PM). Table 2 indicates all the possible iso-PM ROC curves that can be computed with this version of the tool.

Note that, while in the original paper we refer to Matthews Correlation Coefficient as ϕ, this tool uses the term MCC, which should be easier to recognize for people from different research fields.

Table 1: Confusion Matrix.

	Actual Positives	Actual Negatives	
Estimated Positives	TP	FP	EP=TP+FP
Estimated Negatives	FN	TN	EN=TN+FN
	AP=TP+FN	AN=TN+FP	n=AP+AN=EP+EN

Table 2: List of all PMs supported by the tool

Metric	Aliases	Formula	Values Range
TPR	True Positive Rate, Recall	$\frac{TP}{AP}$	[0,1]
FPR	False Positive Rate, Fallout	$\frac{FP}{AN}$	[0,1]
MCC	Matthews Correlation Coefficient, ϕ	$\frac{TP*TN-FP*FN}{\sqrt{EN*EP*AN*AP}}$	[-1,1]
BA	Balanced Accuracy	$\frac{TPR+TNR}{2}$	[0,1]
Gmean	-	$\sqrt{TPR*TNR}$	[0,1]
GM	-	$\frac{2*TPR*TNR}{TPR+TNR}$	[0,1]
D2H	Distance to Heaven (normalized)	$\sqrt{\frac{(1-TPR)^2+FPR^2}{2}}$	[0,1]
precision	PPV, Positive Precision Value	$\frac{TP}{EP}$	[0,1]
F-score	F1	$\frac{2*precision*TPR}{precision+TPR}$	[0,1]
NPV	Negative Precision Value	$\frac{TN}{EN}$	[0,1]
TNR	True Negative Rate, Specificity	$\frac{TN}{AN}$	[0,1]
NM	Negative F1, Negative F-Score	$\frac{2*NPV*TNR}{NPV+TNR}$	[0,1]
Markedness	-	$precision - (1 - NPV)$	[-1,1]

2.1 Cost

This tool also allows to draw families of iso-cost curves. In this tool, "cost" is the misclassification cost, which is computed as

$$MC = cFP * FP + cFN * FN \tag{1}$$

where FP and FN are False Positives and Negatives, respectively, and cFP and cFN are the relative cost for each misclassified element. This tool considers a normalized version of MC, as in Formula 2.

$$NC = \lambda * \frac{1}{1+k} * (1 - TPR) + (1 - \lambda) * \frac{k}{1+k} * FPR \tag{2}$$

Where $\lambda = \frac{cFN}{cFN+cFP}$, and $k = \frac{AN}{AP}$. More details can be found in the following paper [1].

3 How to Use the Tool

Figure 1 shows the tool when opened. On the top-left corner there are four buttons:

- **Menu** Allows player to add new curves, or modify curves that were already loaded in the tool. See Section 3.1 for more details.

- **Open Configuration** Allows the user to open a project saved in a previous session. See Section 3.2 for more details.

- **Reset** Clears the tool, allowing the user to start again. The tool will ask the user to press the "Reset" button again to confirm the choice.

- **Help** Pressing this button will open this PDF file.

Figure 1: First screen of the tool.

Calculate ROC Curve

Menu   Open Project   Reset
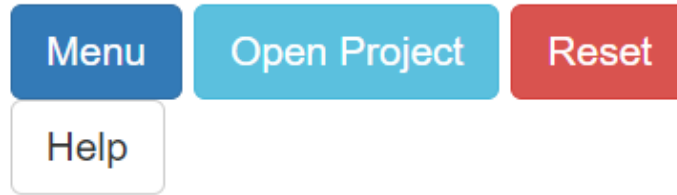
Help

3.1 Using the Menu

Figure 2 shows the menu. In this menu the user can load a ROC curve and decide to generate different iso-PM curves.

Figure 2: Menu.

3.1.1   Loading a ROC Curve

The user can upload a file that represents a ROC curve. In this version, the file must be a .csv file containing at least two columns: one for the FPR values of all the points of the ROC curve, and one for the TPR values. The two columns must be named **FPR** and **TPR**, respectively.

The .csv can have two optional columns as well.

One must be named **Thresholds**, and contains the threshold used to obtain a specific point of the ROC curve.

To upload multiple curves with one file, the user must add a **Name** column. Each curve will have a different name, which has to be assigned to each point of the ROC curve. This column can be used for just one curve as well. If no Name column is present, the tool will assign the name of the file to the curve. It is not possible to load multiple curves with one file without a Name column.

Table 3.1.1 shows an example of .csv, and Figure 3 is the resulting plot.

Table 3: Example of .csv file with multiple curve.

FPR	TPR	Name
0	0	Curve Test 1
0.2	0.3	Curve Test 1
0.7	0.9	Curve Test 1
1	1	Curve Test 1
0	0	Curve Test 2
0.4	0.4	Curve Test 2
0.6	1	Curve Test 2
1	1	Curve Test 2

Before generating the plot, the user needs to set the AP and AN values. These values represent the number of positive (Actual Positives, AP) and negative (Actual Negatives, AN) elements in the test dataset used to test the model and generate the ROC curve. This value can be changed later, but all ROC curves share the same values. This is because, in order to properly confront multiple ROC curves, all curves should be generated from the same test dataset.

The user can then select the size of the plot.

To plot the curve, the user needs to press the "Submit" button.

The result is presented in Figure 4. By default, the plot shows the ROC curve, the boundaries of the RoI (That are, the lines $FPR = \rho$ and $TPR = \rho$, where $\rho = AP/(AP + AN)$), the RoI region, and two points. These two points represent the first and last point of the ROC curve that is inside the RoI.

The textbox below reports some key values for each curve. When plotting just the ROC curve, the textbox shows the AUC, RRA (The area under the curve inside the RoI, divided by the area of the RoI), and the values for the two highlighted points inside the RoI. If thresholds values are provided by the presence of a Thresholds column in the .csv file of the curve, the threshold value of the two points is also reported. If none of the points of the ROC curve is inside the RoI, no point will be highlighted in the plot.

Figure 3: Resulting plot of the .csv file in Table 3.1.1.
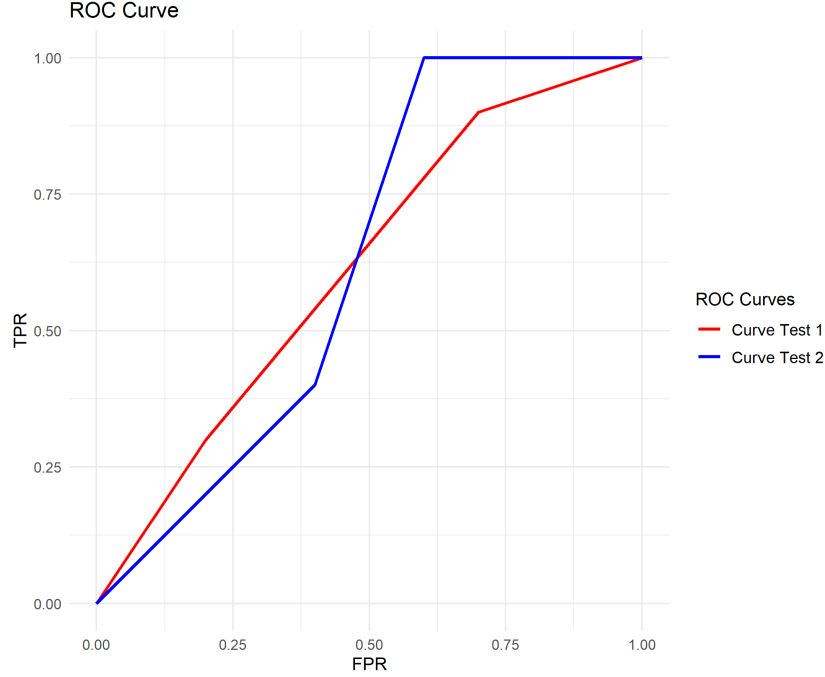
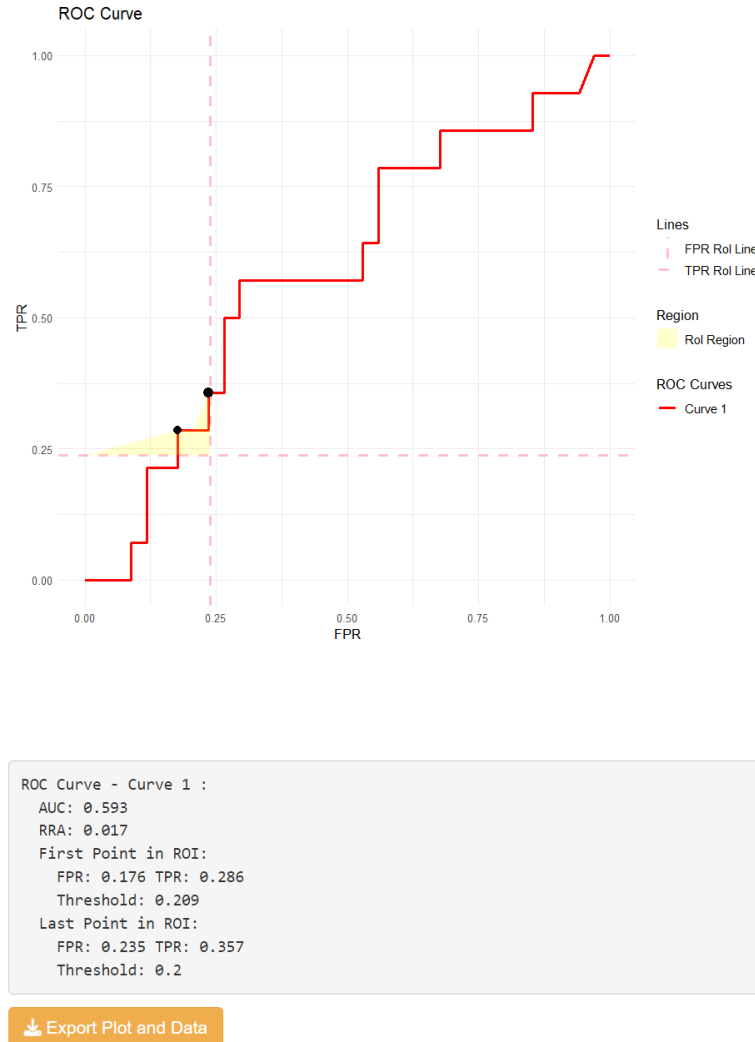### 3.1.2 Doing More Advanced Analysis With Iso-PM ROC Curves

ROC Tool is not just a simple ROC curve plotter. The main feature of the tool is to provide an easy and straight-forward way to plot and study iso-PM ROC curves.

By pressing the "Iso-Curves" button, the relative sub-menu becomes visible (Figure 5). This version of the tool is able to generate a family of iso-PM curves, from the minimum to the maximum values with a step that can be set up in the "Interval Iso-Curve:" textbox. This family of curves allows the user to give a quick glance at which PM interval a specific point of the ROC curve resides in, as seen in Figure 6. To choose which metrics to plot, the user can select a metric from the ones available (see Section 2) in the "Iso-Curve Metric:" textbox. It is also possible to plot iso-PM curves for multiple metrics if needed. In Figure 6 the iso-MCC curves with a interval of 0.2 have been plotted.

3.1.3 Specific Iso-PM Curves with the same AUC and RRA Values

The tools also allows to generate iso-PM curves that have the same AUC and RRA values. This way, the user can obtain a "general" PM value for the entire ROC curve. To do so, the user needs to select the PMs in the "Metric for Iso-Curve based on AUC:" and "Metric for Iso-Curve based on RRA:" (Figure 5)

8

Figure 4: An example of result after pressing "Submit".


```
ROC Curve - Curve 1 :
  AUC: 0.593
  RRA: 0.017
  First Point in ROI:
    FPR: 0.176 TPR: 0.286
    Threshold: 0.209
  Last Point in ROI:
    FPR: 0.235 TPR: 0.357
    Threshold: 0.2
```

⬇ Export Plot and Data

textboxes for AUC and RRA, respectively.

The tool will compute the PM value that generates the iso-PM curve with the same AUC/RRA as the original ROC curves. The resulting iso-PM curves are plotted, and the PM values for each curves are printed in the textbox below the plot. An example is represented in Figure 7.

9

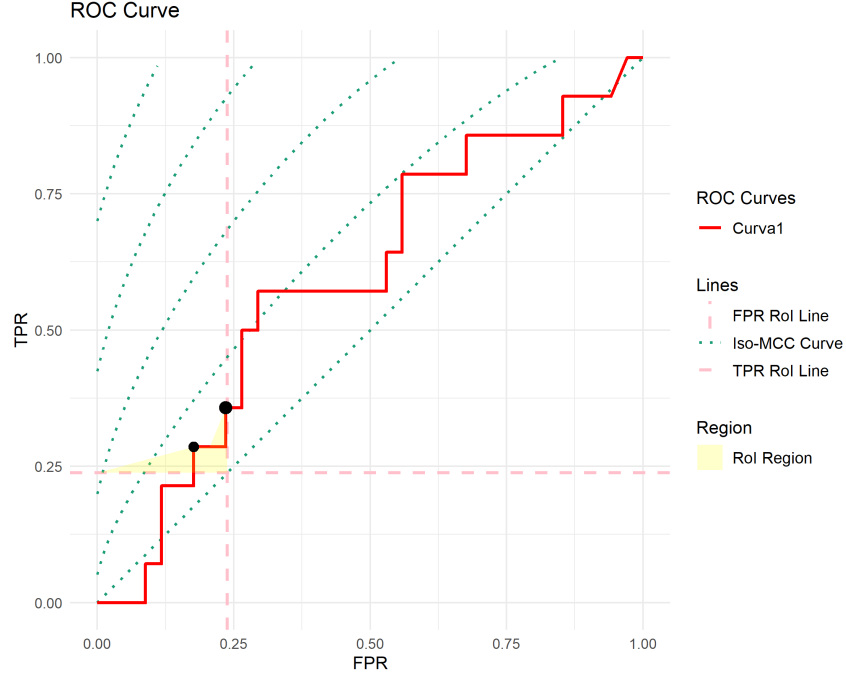Figure 5: Sub-menu that allows the user to create iso-PM curves.

3.1.4 Iso-Cost ROC Curves

To build families of iso-cost curves, the user needs to use a different sub-menu, which is shown by pressing the "Cost" button in the Menu, shown in Figure 8.

Figure 6: Plot showing the curve in Figure 4 with a family of iso-MCC curves with an interval of 0.2.

Unlike other iso-PM curves, the user needs to set the unitary costs for False Positives and Negatives (cFP and cFN). Aside from that, it works exactly like the generation of other iso-PM curves.

### 3.1.5 Updating the Plot, Hide Elements, and Rename the Curve

After pressing the "Submit" button, it is possible to update the values in the Menu. When values are changed, the plot and textbox are automatically updated. There is no need to press "Submit" again, unless the user wants to add a new curve to the plot.

By pressing the "Rename ROC Curves" button, it is possible to change the name assigned to each curve. To apply the changes, the user needs to press the "Save Changes" button.

By pressing the "Show/Hide Lines" button, a list of checkboxes is shown. These checkboxes indicate what to plot and what to hide. This allows users to load many ROC curves, generating multiple iso-PM curves, and making multiple analysis, without having to reset, and without making the plot completely unreadable.

The names are self-explanatory.

Figure 7: Iso-MCC curves added for two ROC curves.
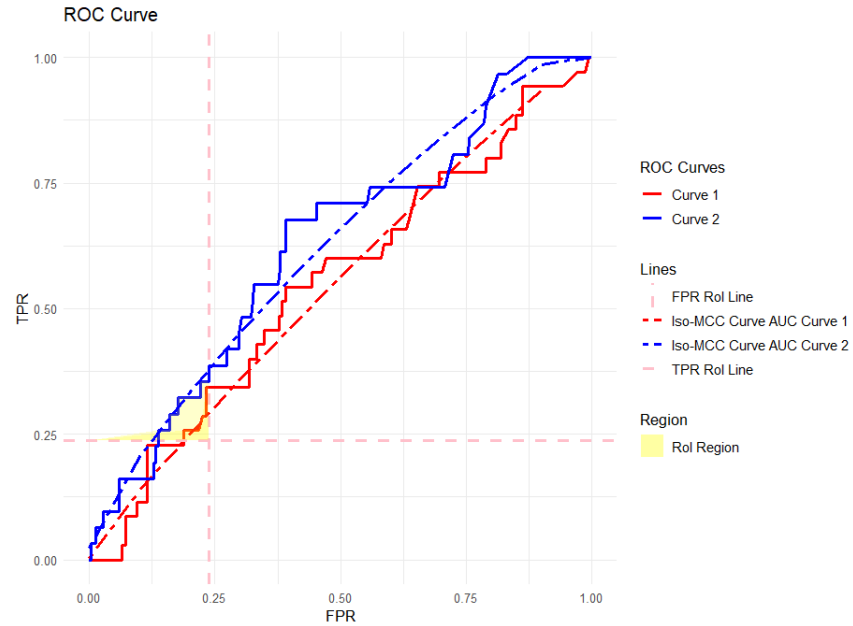

```
ROC Curve - Curve 1 :
  AUC: 0.549
  RRA: 0.01
    Iso-MCC AUC:0.053
  First Point in ROI:
    FPR: 0.188 TPR: 0.257
  Last Point in ROI:
    FPR: 0.232 TPR: 0.343

ROC Curve - Curve 2 :
  AUC: 0.625
  RRA: 0.039
    Iso-MCC AUC:0.136
  First Point in ROI:
    FPR: 0.138 TPR: 0.258
  Last Point in ROI:
    FPR: 0.236 TPR: 0.355
```

⬇ Export Plot and Data

3.2   Save and Load a Project

It is possible to save progress to load them at a later time, or to share them
with other users.

Figure 8: Iso-cost sub-menu.

Cost

FN (cFN):

1

FP (cFP):

1

Interval Iso-Curves Cost:

0,1

To do so, the user needs to press the "Save Project" button. Now, the user can choose where to save the project and rename it. The project is saved in a .rds file.

To load a project, the user needs to press the "Open Project" button, next to the "Menu" button and select the .rds file. The tool will update all values in the menu's fields, generate the plot, and print the results.

4 Exporting the Results

If the user wants to export the plot and results obtained, they can press the "Export Plot and Data" button below the textbox (Figure 4). The user then needs to choose a location a rename the file.

The exported file is a compressed .zip folder. This folder contains two files: `roc_data.xhtml` and `roc_plot.png`. By extracting the folder, it is possible to open the .xhtml file to see all the produced data and the plot. The .png file only contains the plot.

Figure 9: The resulting .xhtml file of an exported project. The plot is slightly cut in the Figure to make it more readable in this Documentation.

Curve Name: Curve1

AUC: 0.549

RRA: 0.01

First point in RoI: FPR = 0.188, TPR = 0.257

Last point in RoI: FPR = 0.232, TPR = 0.343

Iso-MCC AUC:0.053
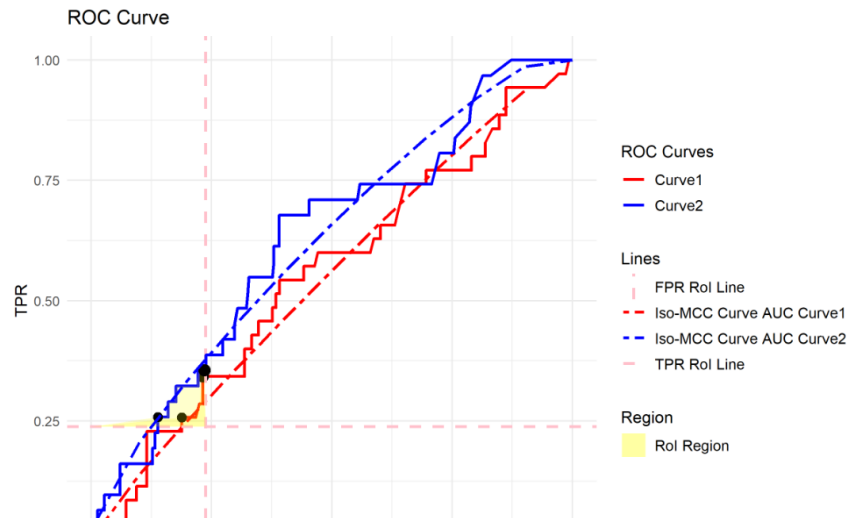
Curve Name: Curve2

AUC: 0.625

RRA: 0.039

First point in RoI: FPR = 0.138, TPR = 0.258

Last point in RoI: FPR = 0.236, TPR = 0.355

Iso-MCC AUC:0.136

Show/Hide Plot

References

[1] Sandro Morasca and Luigi Lavazza. On the assessment of software defect prediction models via roc curves. *Empirical Software Engineering*, 25(5):3977–4019, 2020.