Performance metrics

AI For You

Summary

- **I/ Performance metrics for regression**
- II/ Performance metrics for classification

1. Notations

- N – Number of observations
- $y \in \mathbb{R}^N$ - Target value
- $\hat{y} \in \mathbb{R}^N$- prediction
- $\hat{y}_i \in \mathbb{R}$ - prediction for the i-th observation
- $y_i \in \mathbb{R}$ - real value for the i-th observation

2. Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

- The most popular metric for measuring regression error
- The error is difficult to interpret because we measure the square of the error

3. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

- Each minimizer for MSE is a minimizer for RMSE
- The error scale is the same as the target scale
- But it is a bit easier to work with MSE

4. R-squared

$$R^2 = 1 - \frac{MSE}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- Optimization for RMSE and MSE is also an optimization for R-square
- With RMSE and MSE, it is difficult to estimate if our model is good enough
- R-square is between 0 (poor model) and 1 (perfect model)

5. Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

- RMSE, MSE, R-squared penalize large errors more
- MAE is less sensitive to outliers than R-squared, MSE and RMSE

6. Mean Square Percentage Error

$$MSPE = \frac{100\%}{N} \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2$$

Predicted	Sold	MSE	MSPE
9	10	1	1
999	1000	1	0,0001

7. Mean Absolute Percentage Error

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Predicted	Sold	MAE	MAPE
9	10	1	10
999	1000	1	0,1

8. Let's sum up

- MSE, RMSE, R-squared
  - They are the same from the point of view of optimization
- MAE
  - Robust to outliers
- (R)MSPE
  - Weighted version of the (R)MSE
- MAPE
  - Weighted version of the MAE

Summary

- I/ Performance metrics for regression
- II/ Performance metrics for classification

1. Notations

- N – number of observations
- L – number of classes
- y - ground truth
- $\hat{y}$ - predictions
- $[a = b]$ – indicator function
- ’soft labels’ (soft predictions) – classifier’s scores
- ‘hard labels’ (hard predictions) - classifier’s Boolean
    - $arg \max_i f_i(x)$
    - [f(x) > b], b-threshold

2. Accuracy score

Accuracy = 0.99 !!

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N}[\hat{y}_i = y_i]$$

True labels	Predictions
10 cats	0 cats
990 dogs	1000 dogs

- Need hard prediction
- How frequent our class prediction is correct.
- Accuracy between 0 and 1, the higher the better.

3. Cross-entropy loss or log loss

Binary:

$$Logloss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

Multiclass:

$$Logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{L} y_{il}\log(\hat{y}_{il})$$

4. Confusion matrix

		Predicted label	
		$\hat{y} = 0$	$\hat{y} = 1$
True label	$y = 0$	6 TN	9 FP
	$y = 1$	1 FN	10 TP

Specificity = $\frac{TN}{TN+FP}$ = 1 - FPR

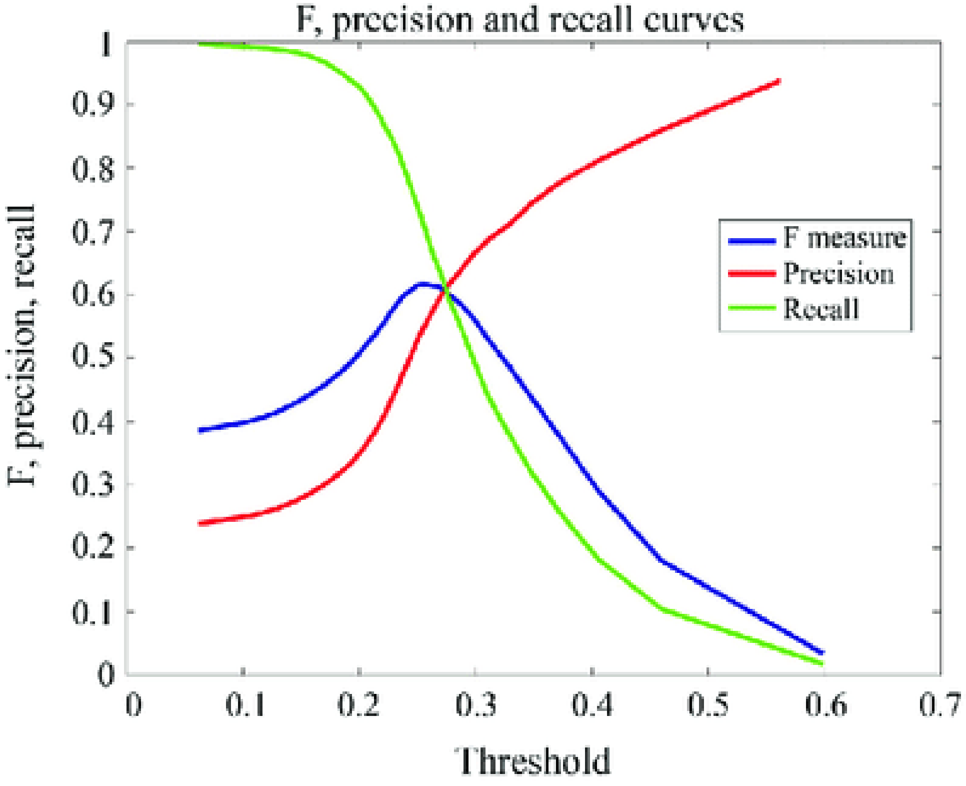
FPR = $\frac{FP}{FP+TN}$ = 1 - Specificity

Sensitivity, Recall = $\frac{TP}{TP+FN}$ = TPR

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

FP = False Positive
FN = False Negative
TP = True Positive
TN = True Negative
FPR = False Positive Rate
TPR = True Positive Rate

5. Precision, Recall, F score



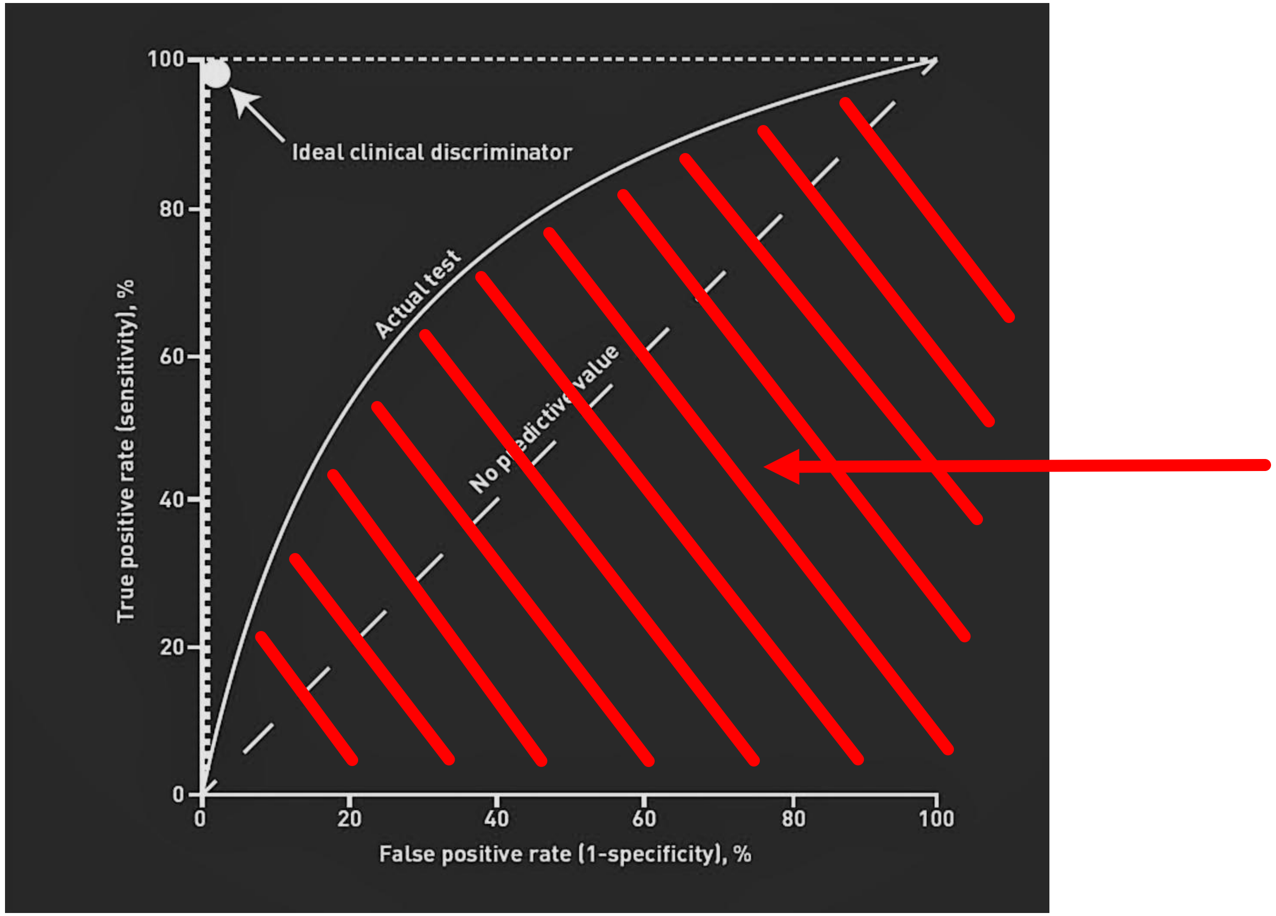
Recall, Sensitivity = $\frac{TP}{TP+FN}$

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

$$F_\beta = (1+\beta)\frac{precision\,.\,recall}{\beta^2.\ precision + recall}$$

$$F_1 = F = 2\ \frac{precision\,.\,recall}{precision + recall}$$

6. AUC & ROC curve



Area Under Curve (AUC)