Mikhail Hushchyn

Quality Metrics

Classification and Regression

2021

NATIONAL RESEARCH UNIVERSITY
LAMBDA • HSE
Yandex
SCHOOL OF DATA ANALYSIS
EPFL
Schaffhausen Institute of Technology

Outline

- Quality metrics for regression
- Quality metrics for classification

Quality Metrics for Regression

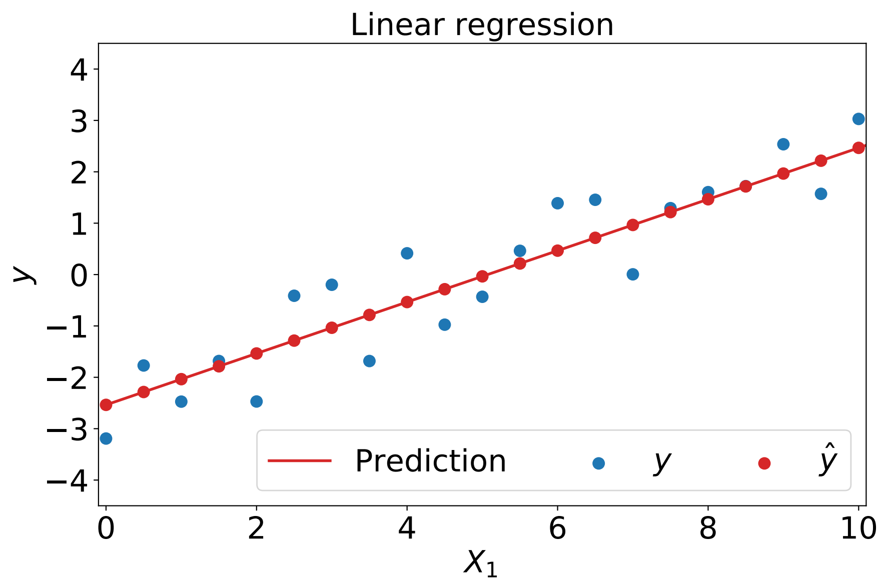
Problem formulation

Consider a dataset X, y and a linear regression model:

$$\hat{y} = Xw$$

where w – weights of the model.

The goal is to measure the quality of this model, estimate how close predictions $\hat{y}$ to the real values y.



Popular quality metrics

- Root Mean Squared Error (RMSE):

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

- Mean Absolute Error (MAE):

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

- It is hard to tell if a model is good: $RMSE = 1$ represents the different quality of a model for $\bar{y} = 100$ and $\bar{y} = 1$

Other quality metrics #1

- Mean Absolute Percentage Error (MAPE):

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

- Measures relative error of the prediction
- Easy to understand quality of the model
- Sensitive to y scale

Other quality metrics #2

- Relative Squared Error (RSE):

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

- Relative Absolute Error (RAE):

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

- RSE shows how the prediction errors differ from the standard deviation of the real values
- Robust to y scale

Other quality metrics #3

- Root Mean Squared Logarithmic Error (RMSLE):

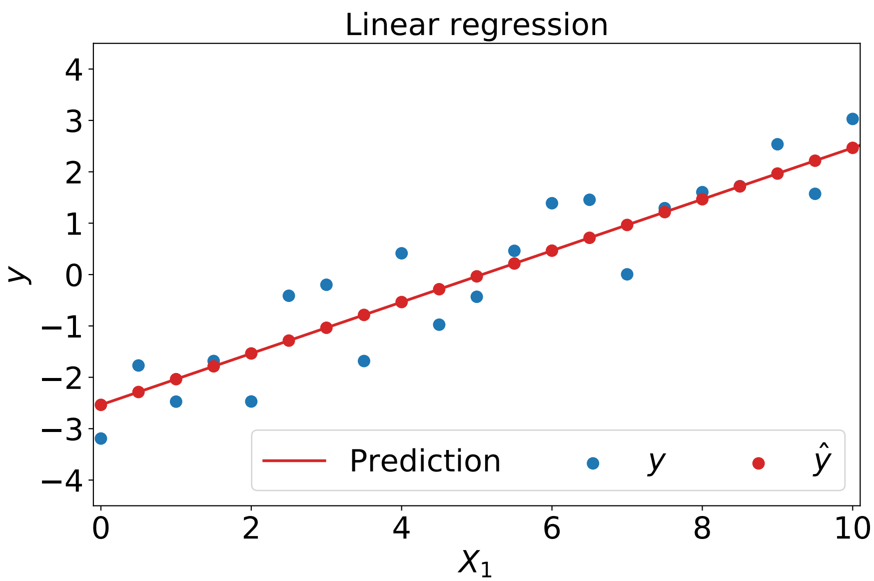
$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(y_i + 1) \ - \log(\hat{y}_i + 1))^2}$$

- It is a great choice, when y_i changes in several orders: $y_i \in [0, 10^6]$

Demonstration

Metric	No outliers
RMSE	0.67
MAE	0.59
MAPE, %	1035
RSE	0.39
RAE	0.40

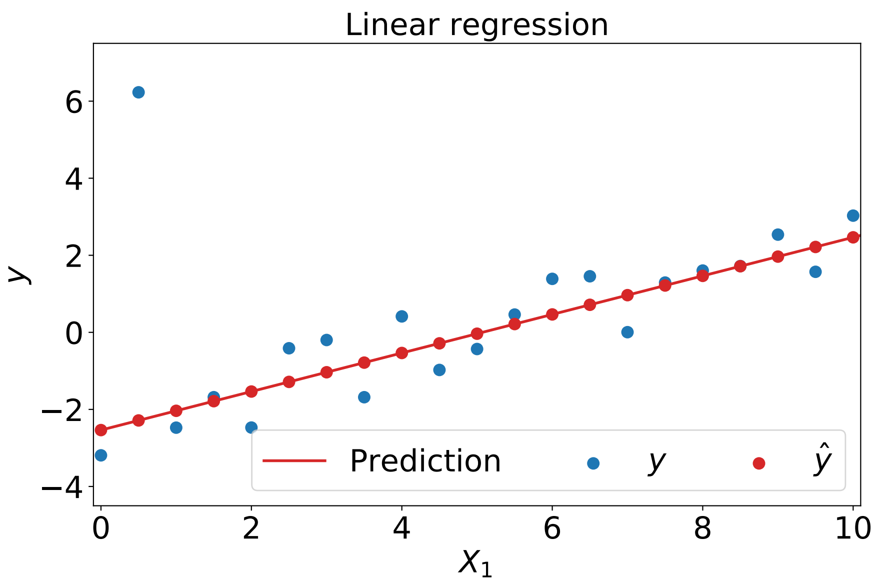
- MAPE fails because of y scale and y_i that are close to 0



Demonstration

Metric	No outliers	With outlier
RMSE	0.67	1.93
MAE	0.59	0.96
MAPE, %	1035	1040
RSE	0.39	0.92
RAE	0.40	0.58

- Outliers significantly affect the metrics
- MAE and RAE are more robust

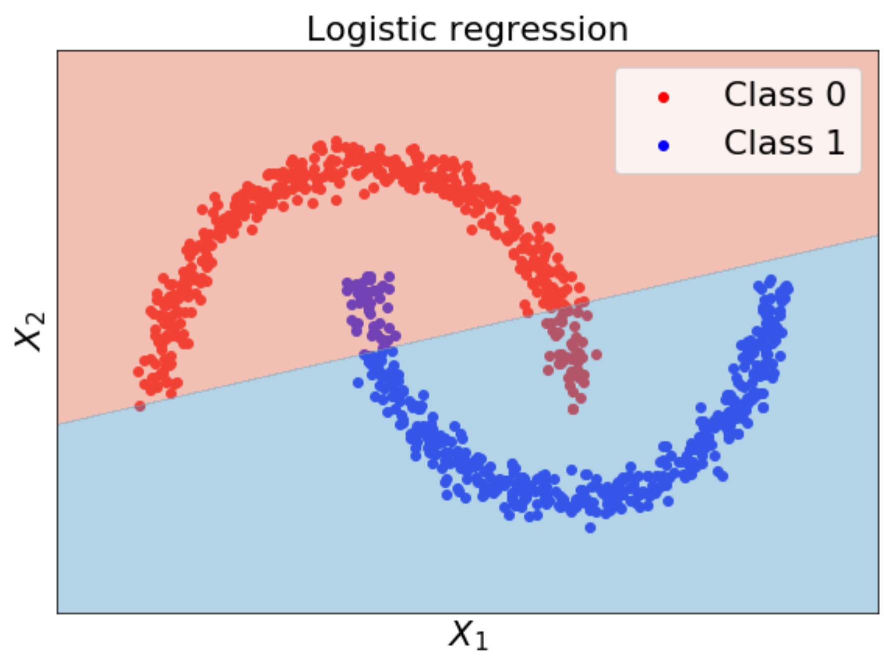


Quality Metrics for Classification

Problem formulation

Consider a binary classification problem with a data sample and a classifier.

The goal is to measure the quality of the classifier, estimate how well it separates objects of different classes.



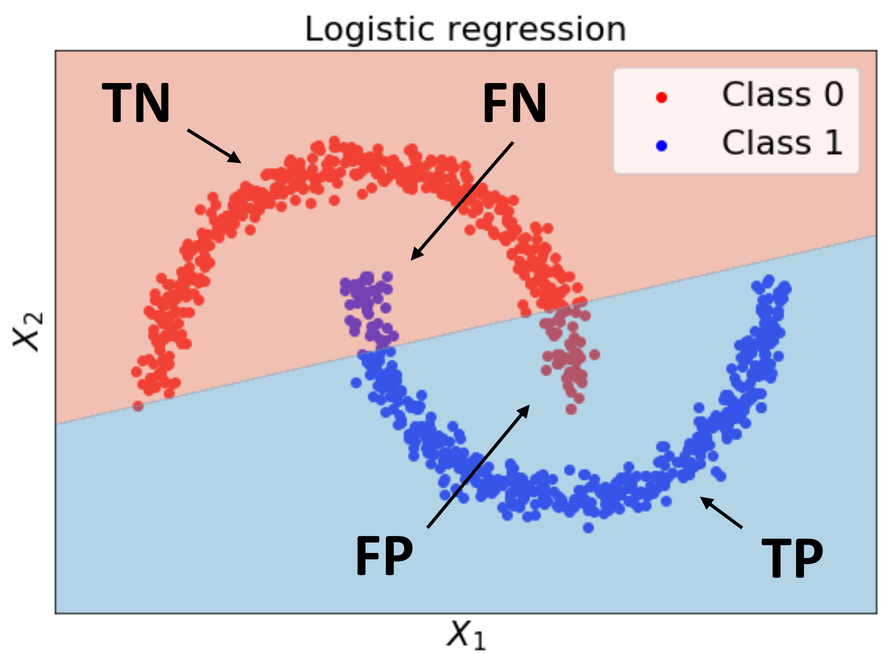
Confusion matrix

- **TP** (True Positive) – correctly predicted positives
- **FP** (False Positive) – predicted as positives, but negatives (1st order error)
- **TN** (True Negative) – correctly predicted negatives
- **FN** (False Negative) – predicted as negatives, but positives (2nd order error)

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Confusion matrix

- **TP** (True Positive) – correctly predicted positives
- **FP** (False Positive) – predicted as positives, but negatives (1^{st} order error)
- **TN** (True Negative) – correctly predicted negatives
- **FN** (False Negative) – predicted as negatives, but positives (2^{nd} order error)



Confusion matrix

- All positives ($\boldsymbol{Pos}$):

$$Pos = TP + FN$$

- All negatives ($\boldsymbol{Neg}$):

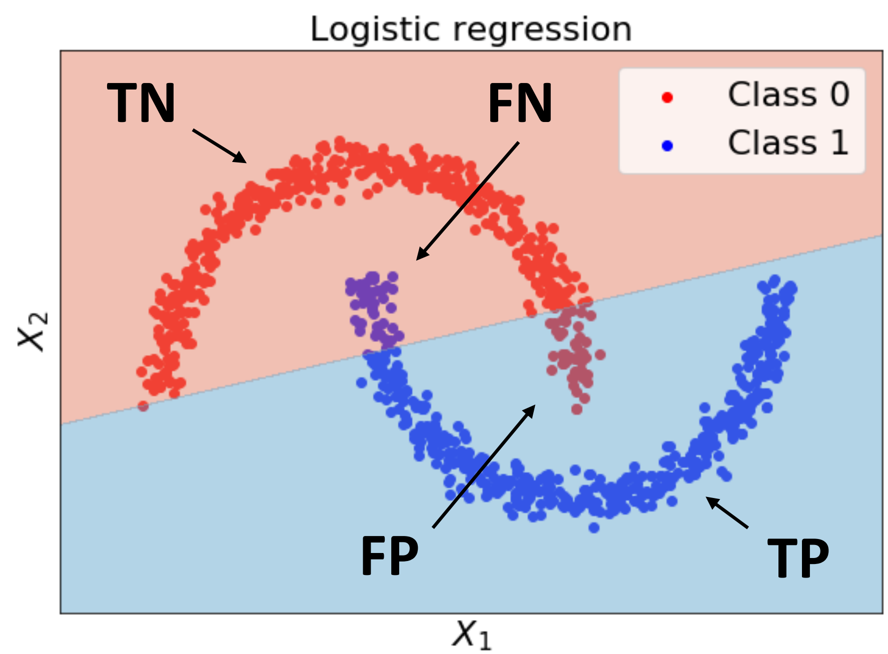
$$Neg = TN + FP$$

- All positive predictions ($\boldsymbol{PosPred}$):

$$PosPred = TP + FP$$

- All negative predictions ($\boldsymbol{NegPred}$):

$$NegPred = TN + FN$$



Quality metrics #1

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} = \frac{TP + TN}{Pos + Neg}$$

- Error rate:

$$\text{Error rate} = 1 - \text{Accuracy}$$

- They measure classification quality for both classes

Quality metrics #2

- Precison:

$$\text{Precison} = \frac{TP}{TP + FP} = \frac{TP}{PosPred}$$

- Recall:

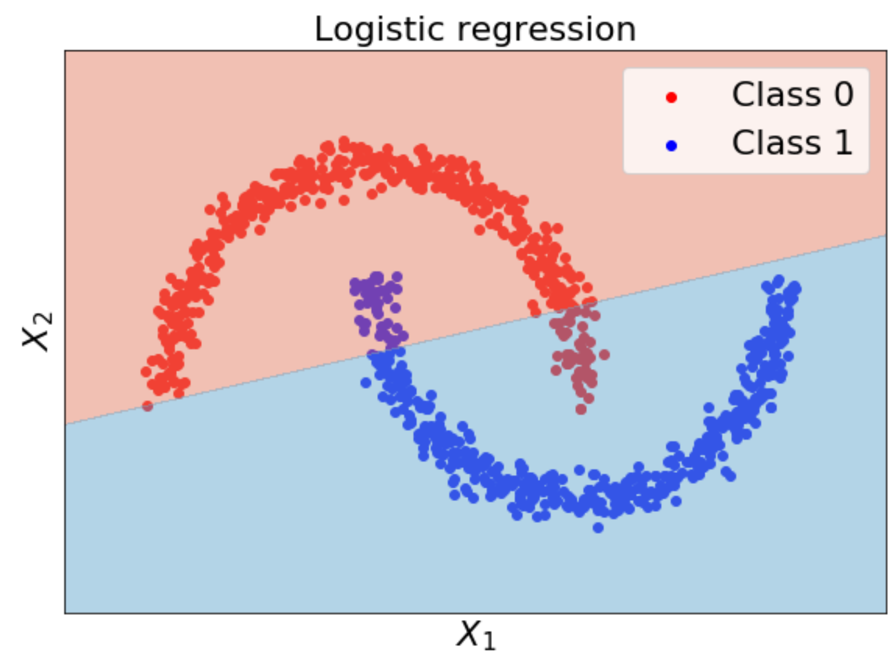
$$\text{Recall} = \frac{TP}{TP + FN} = \frac{TP}{Pos}$$

- F_1-score:

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

Example

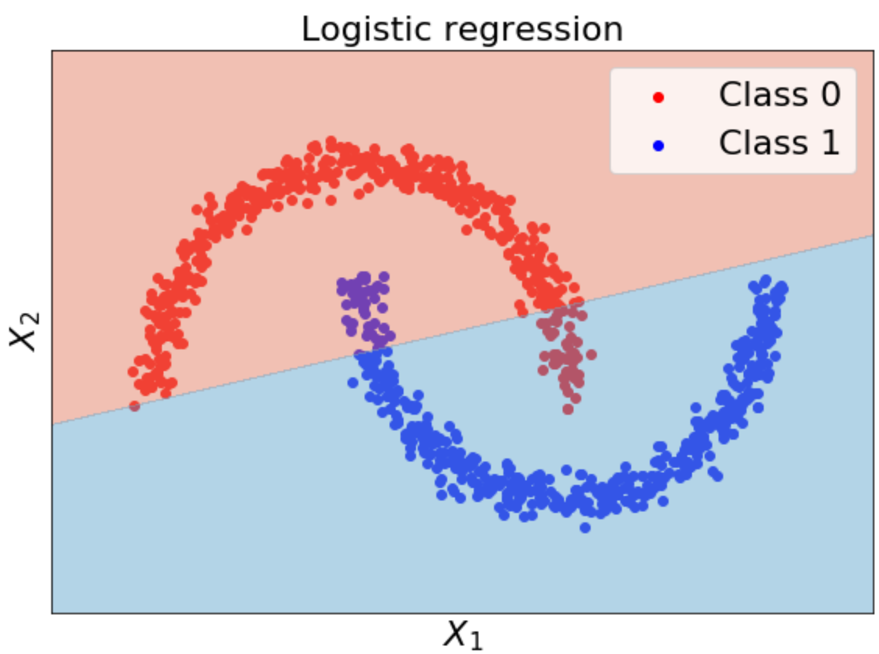
Metric	Value
Accuracy	0.89
Precision	0.89
Recall	0.89
F_1	0.89



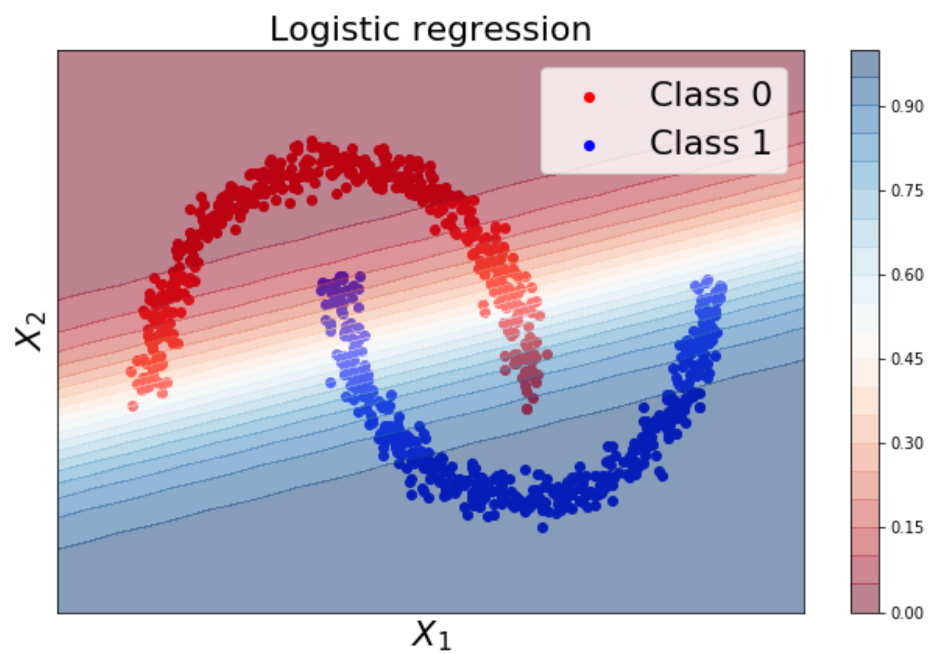
- In this symmetric case values of all metrics are the same
- Latter, we will see other cases

Class label vs class probability

Predict **1** if $\boldsymbol{p} \geq 0.5$
Predict **0** if $\boldsymbol{p} < 0.5$



Probability of positive class $\boldsymbol{p}$:



ROC curve

- ROC (Receiver operating characteristic) curve is a dependency of $\boldsymbol{TPR(\mu)}$ **from** $\boldsymbol{FPR(\mu)}$ for different thresholds μ of the probability of positive class $\boldsymbol{p}$

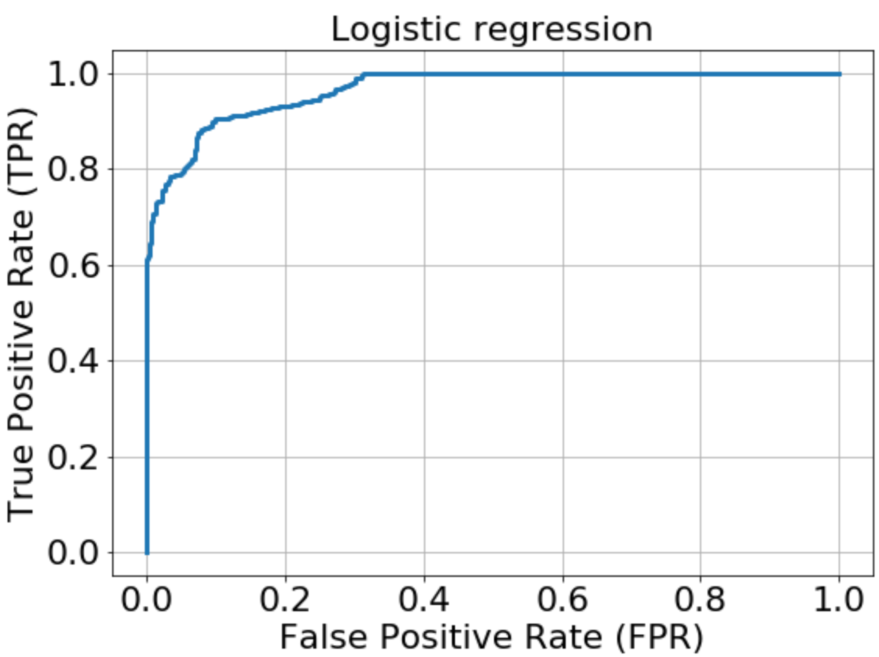
- $TPR(\mu)$ (True Positive Rate):

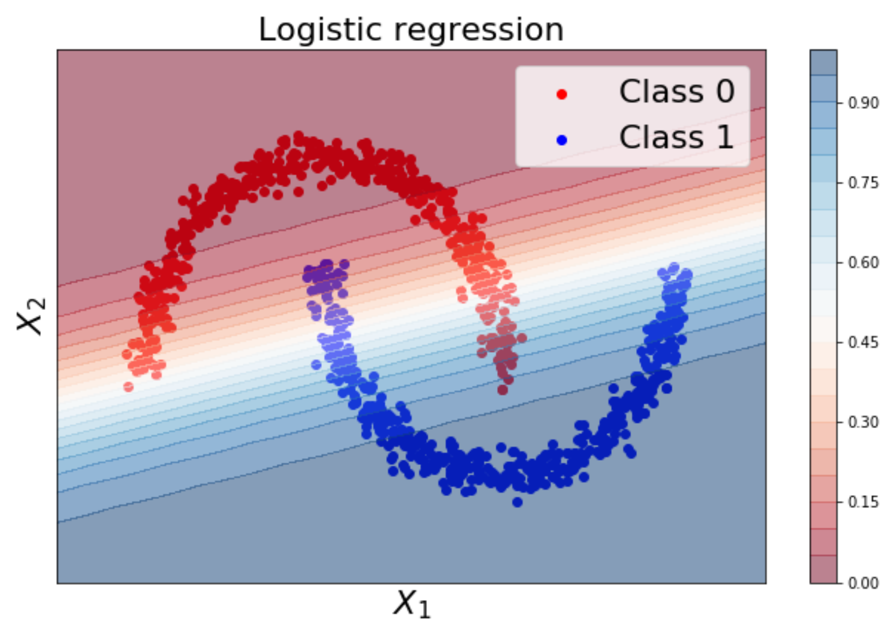
$$TPR(\mu) = \frac{1}{Pos} \sum_{i \in Pos} I[p_i \geq \mu] = \frac{TP(\mu)}{Pos}$$

- $FPR(\mu)$ (False Positive Rate):

$$FPR(\mu) = \frac{1}{Neg} \sum_{i \in Neg} I[p_i \geq \mu] = \frac{FP(\mu)}{Neg}$$

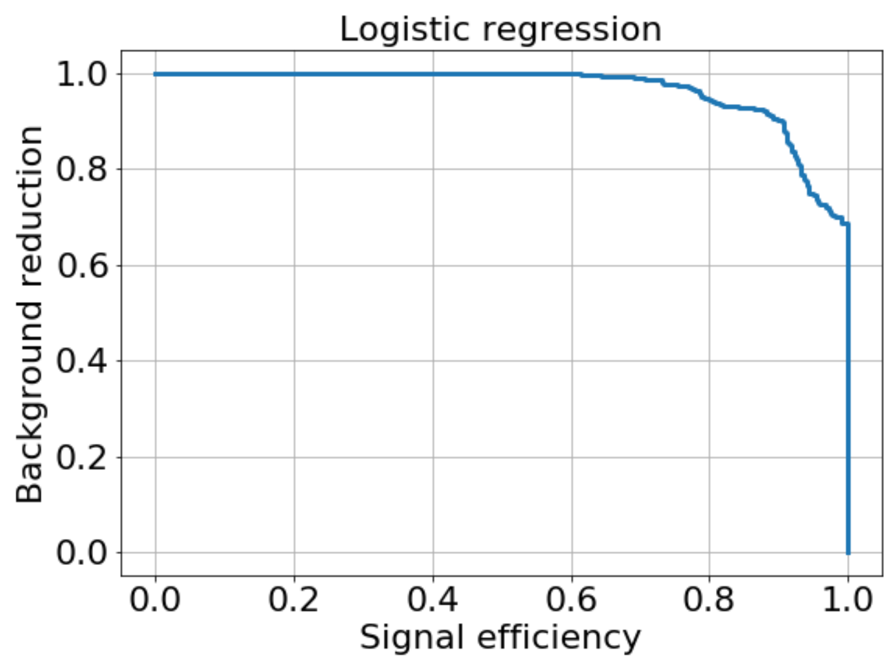
ROC curve


Logistic regression
True Positive Rate (TPR)
False Positive Rate (FPR)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


Logistic regression
Class 0
Class 1
X_2
X_1
0.90
0.75
0.60
0.45
0.30
0.15
0.00

ROC curve

In physics, very often plot dependency of **background reduction** from **signal efficiency**



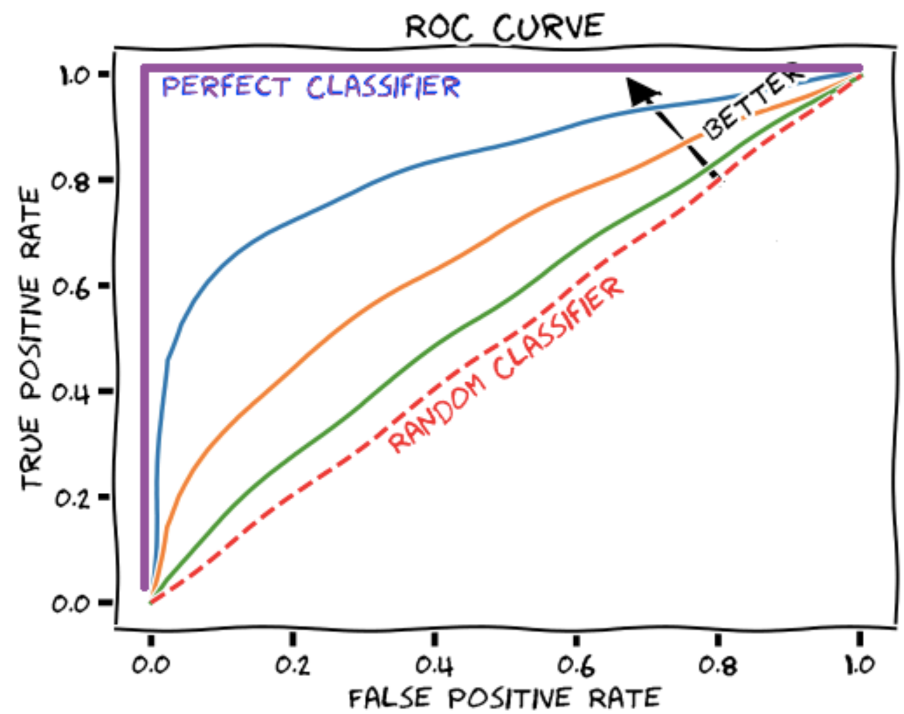
Here:

- **Signal efficiency** = TPR
- **Background reduction** = 1 - FPR

ROC AUC

- ROC curves can be compared using area under the ROC curve (ROC AUC)

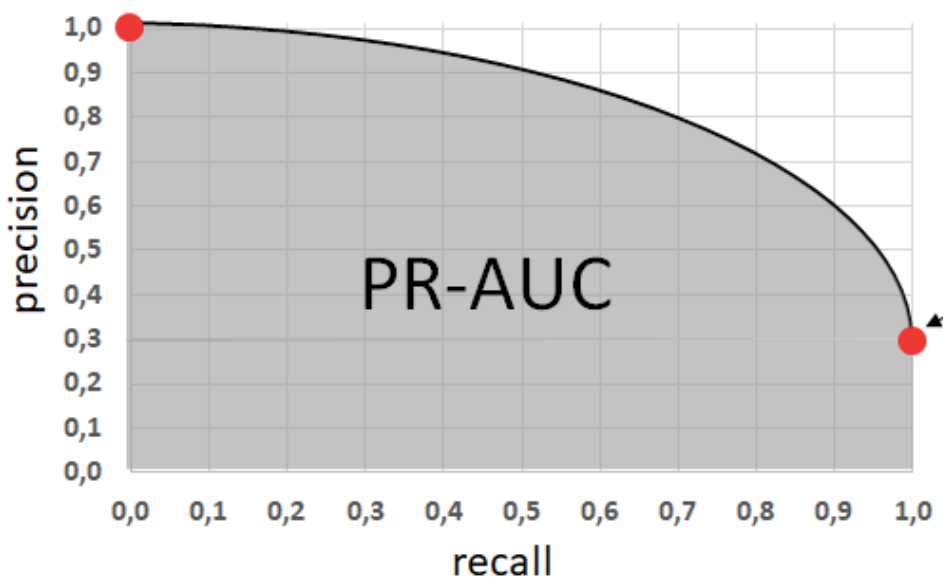
- ROC AUC $\in [0, 1]$ range
- ROC AUC = 0.5 means random guessing
- ROC AUC = 1 means ideal classification
- ROC AUC = 0 also means ideal classification, but for opposite labels ☺



Img: https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/

Precision-Recall curve

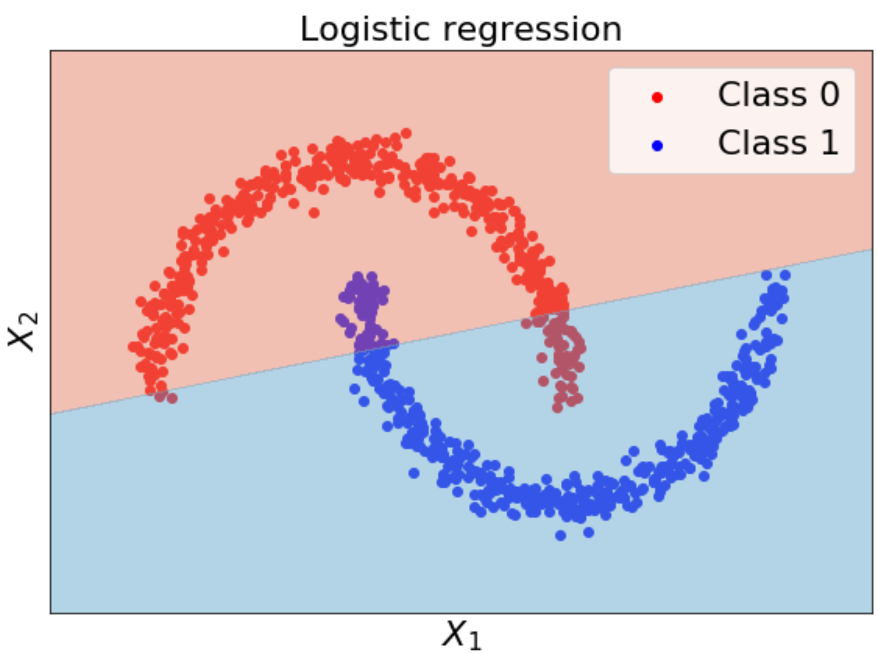
- Similarly to ROC curve, you can plot Precision-Recall curve (PR)
- PR is dependency of **Precision(μ) from Recall(μ)** for different thresholds μ of the positive class probability $\boldsymbol{p}$



Demonstration

Metric	1:1	1:10	10:1
Accuracy	0.89		
Precision	0.89		
Recall	0.89		
F_1	0.89		
ROC AUC	0.97		

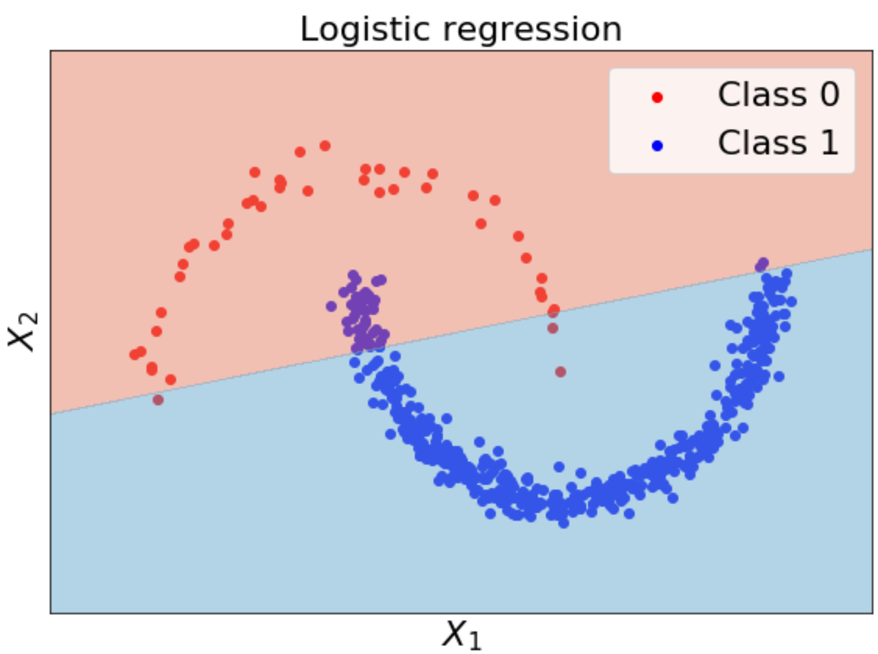
- Let's train a model on a sample with equal number of objects in each class
- **We fix the model** and will change class balance in test sample



Demonstration

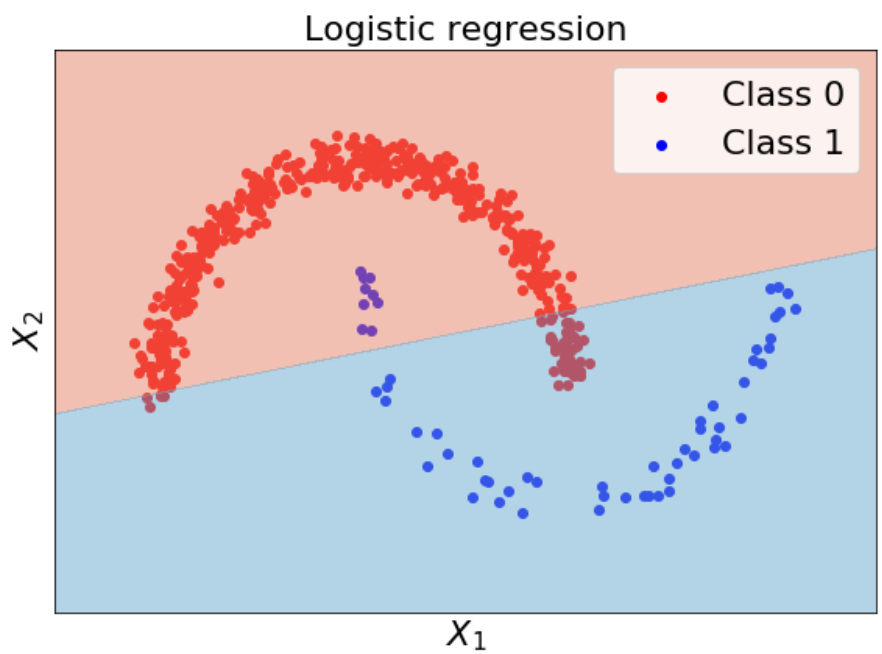
Metric	1:1	1:10	10:1
Accuracy	0.89	0.89	
Precision	0.89	0.99	
Recall	0.89	0.89	
F_1	0.89	0.94	
ROC AUC	0.97	0.97	

- With the class balance changing, some metrics change



Demonstration

Metric	1:1	1:10	10:1
Accuracy	0.89	0.89	0.89
Precision	0.89	0.99	0.47
Recall	0.89	0.89	0.89
F_1	0.89	0.94	0.61
ROC AUC	0.97	0.97	0.97



- **Recall** and **ROC AUC** do not change with the class balance changings
- For Accuracy it is not true in general case

Summary

Summary

- Quality metrics for regression
  - RMSE, MAE, MAPE
  - RSE, RAE, RMSLE
- Quality metrics for classification
  - Confusion matrix
  - Accuracy, precision, recall, F_1-score
  - ROC curve, ROC AUC
  - Precision-Recall curve