Evaluation Metrics for Classification

Example Classification with two Classes

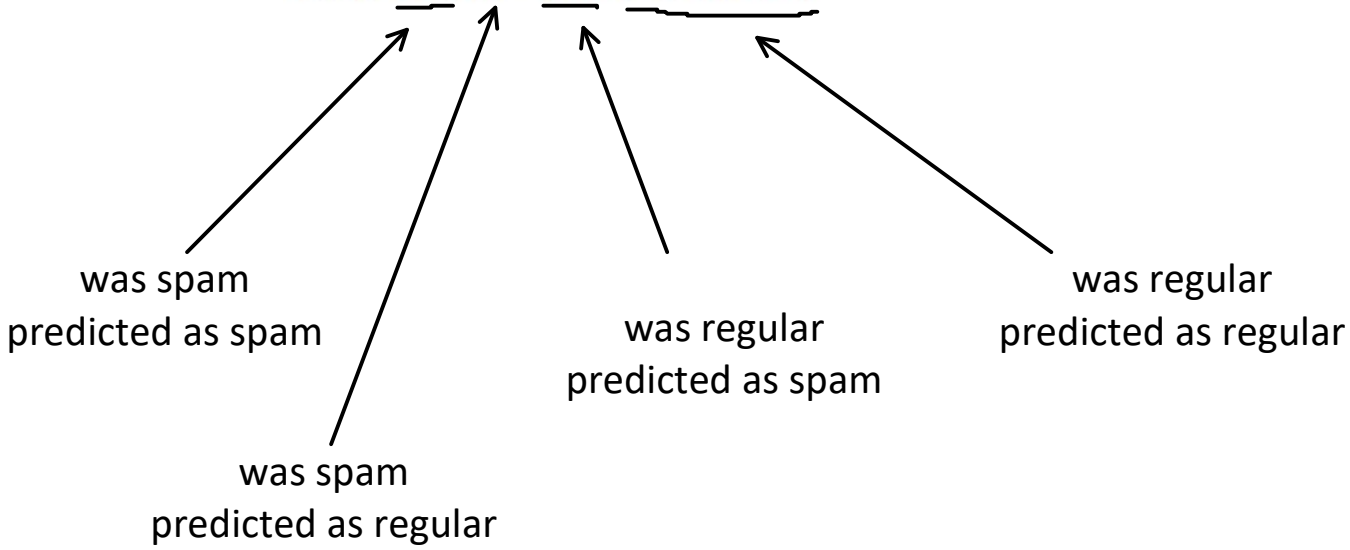
Prof. Dr. Ingo Claßen

Sonntag, 5. Dezember 2021 10:40

Label	Class
1	Spam Mail (positive)
0	Regular Mail (negative)

True vs Predicted

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0

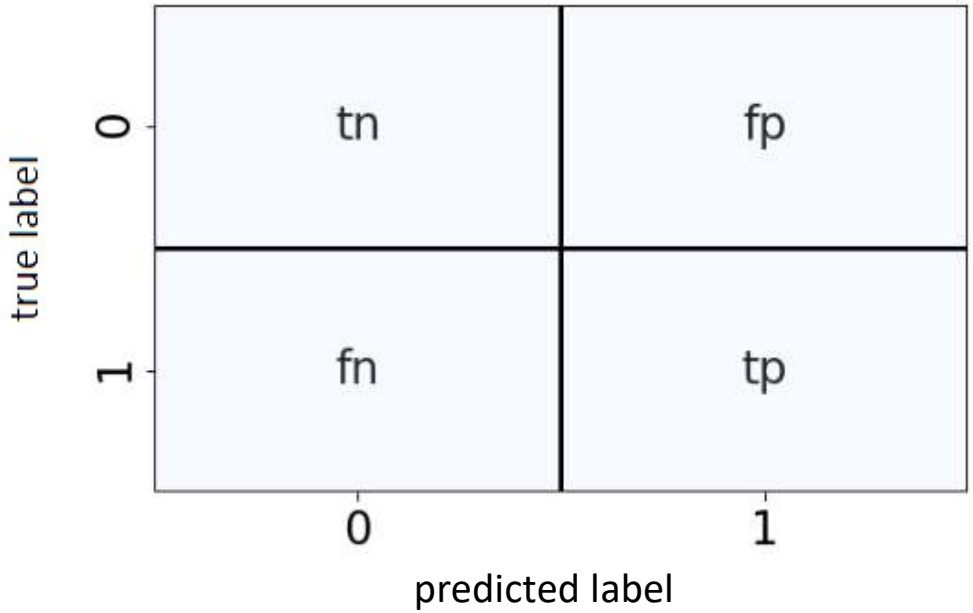


Confusion Matrix (Two Classes)

Prof. Dr. Ingo Claßen

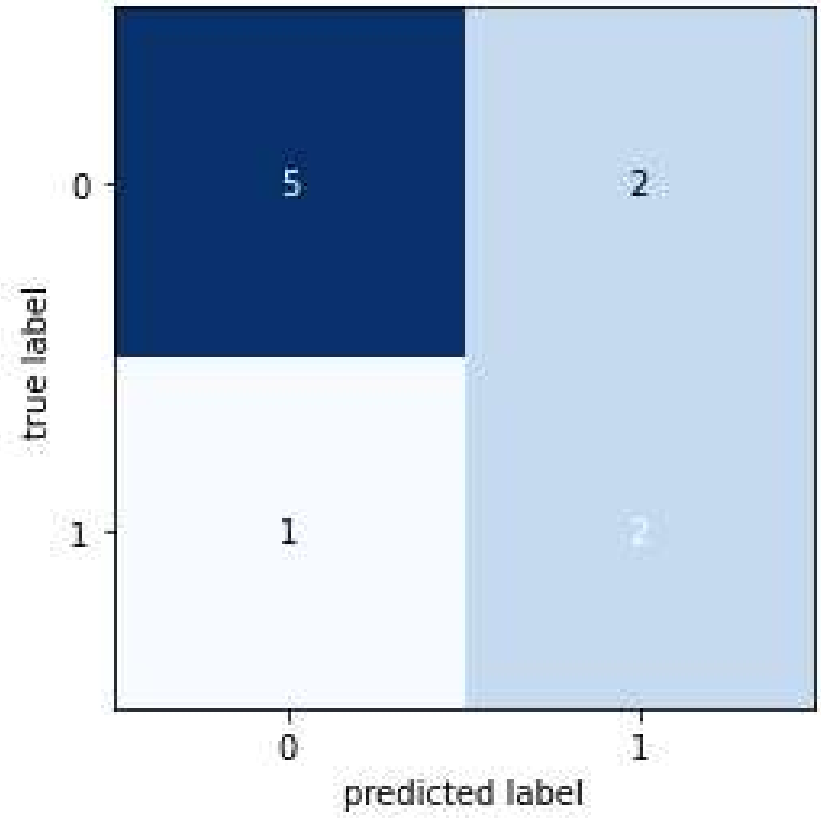
Sonntag, 5. Dezember 2021 10:40

true label	predicted label	name	description
positive	positive	**tp**	true positive predicted
negative	negative	**tn**	true negative predicted
positive	negative	**fn**	false negative predicted
negative	positive	**fp**	false positive predicted



Label	Class
1	positive
0	negative

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0

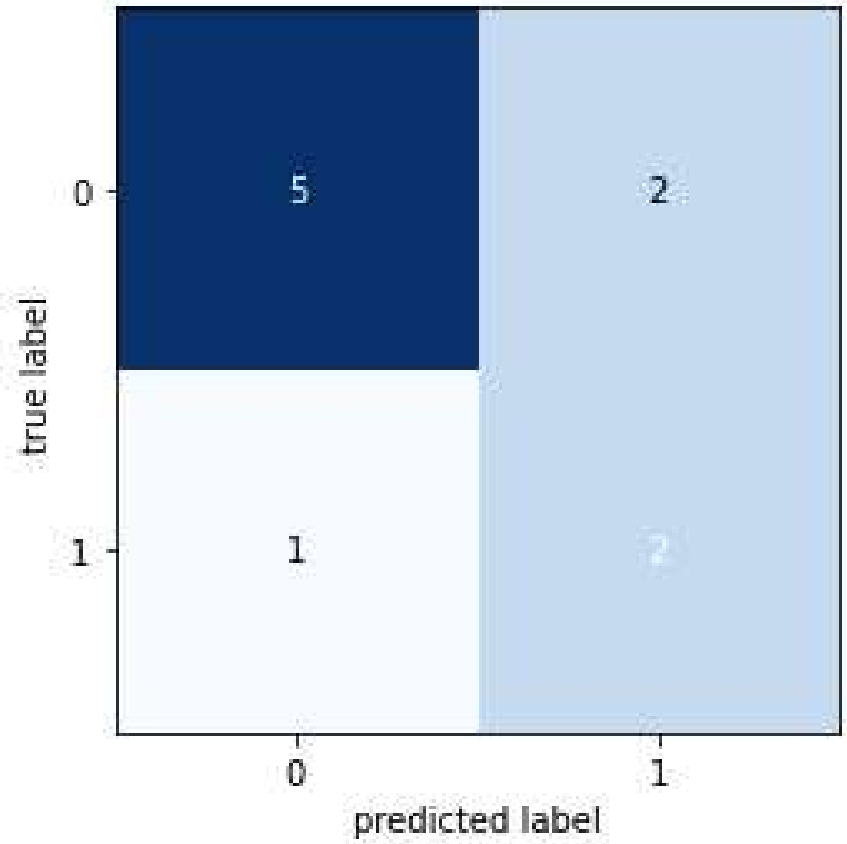


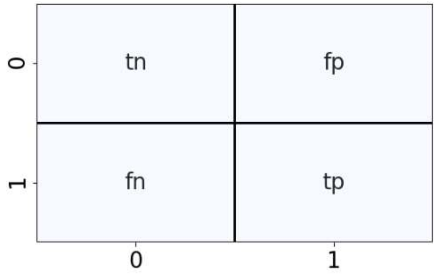
Accuracy

Prof. Dr. Ingo Claßen

Sonntag, 5. Dezember 2021 11:11

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0





$$\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}}$$

$$\text{accuracy} = \frac{2+5}{2+2+5+1} = 0.7 = 70\%$$

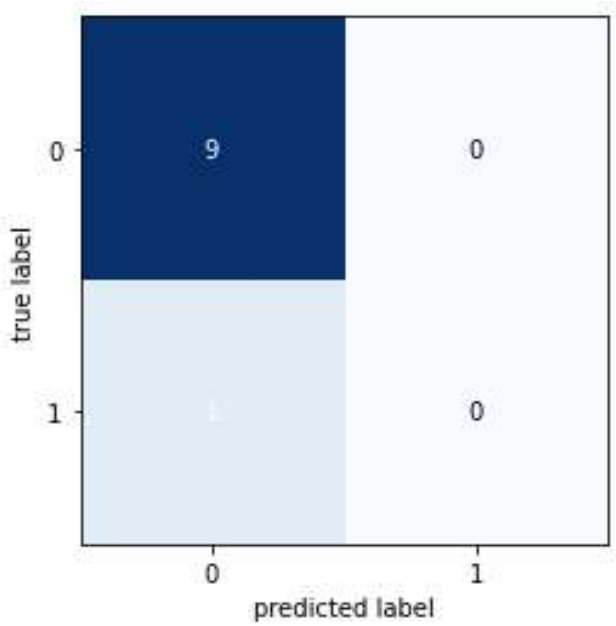
Accuracy - often not a good metrics

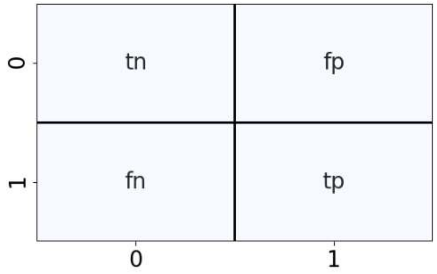
Prof. Dr. Ingo Claßen

Sonntag, 5. Dezember 2021 16:10

- Assumption: Positive cases are rare
- Dummy classificator: Always predicts *negative*

true	1	0	0	0	0	0	0	0	0	0
pred	0	0	0	0	0	0	0	0	0	0





$$\text{accuracy} = \frac{0+9}{0+0+9+1} = 0.9 = 90\%$$

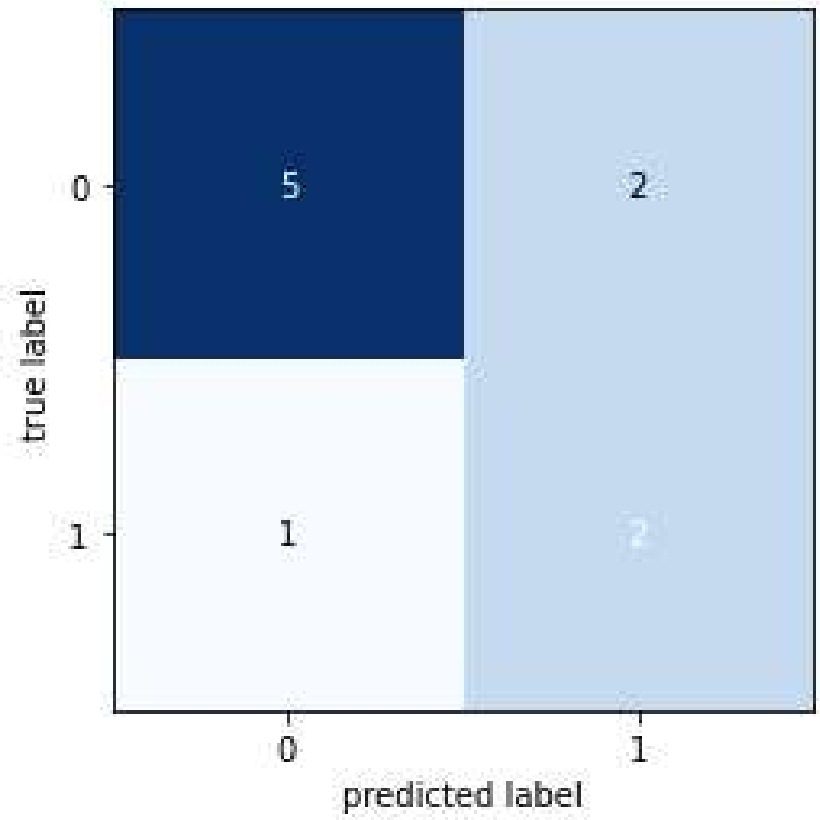
but classifier is useless

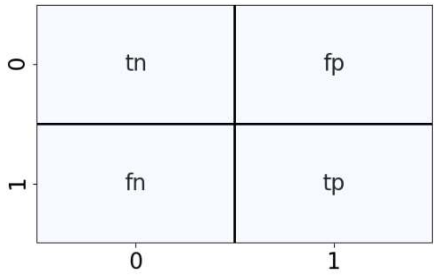
Recall (True Positive Rate - TPR, Hit Rate)

Sonntag, 5. Dezember 2021 16:14

Fraction of true positive predicted to all positives

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0





$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

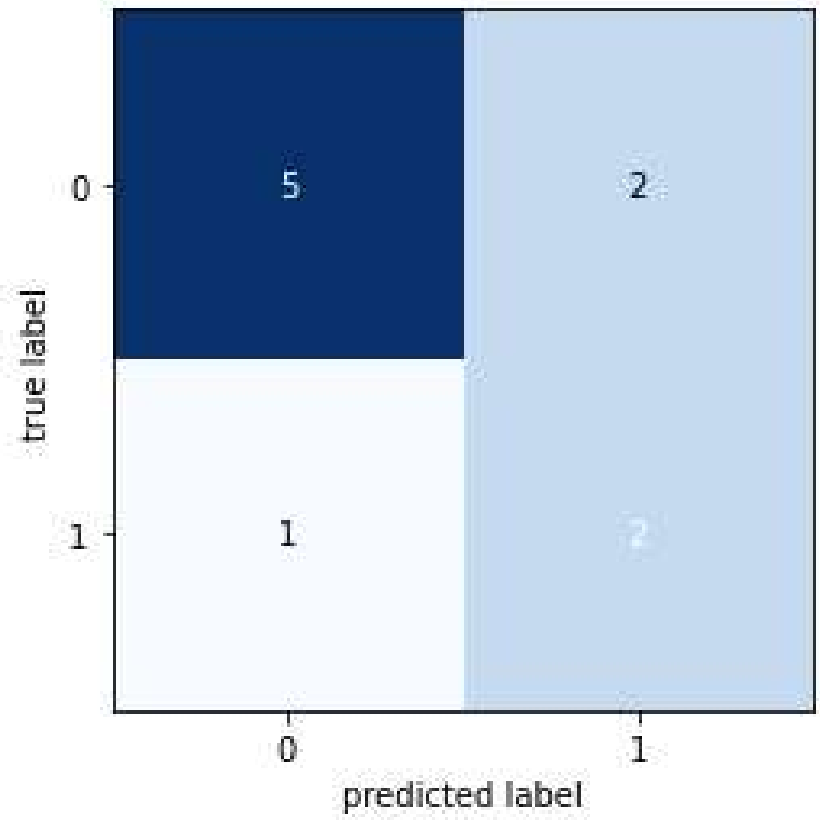
$$\text{recall} = \frac{2}{2+1} = 0.66 = 66\%$$

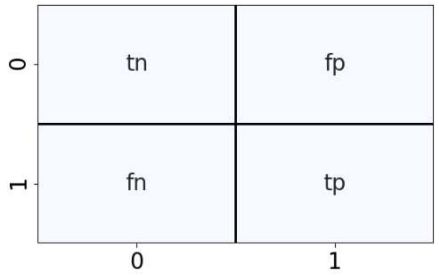
Precision

Sonntag, 5. Dezember 2021 16:14

Fraction of true positive predicted to all positive predicted

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0





$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

$$\text{precision} = \frac{2}{2+2} = 0.5 = 50\%$$

- Easy to get high recall: just predict positive
  - recall = 100% -> precision bad
- Easy to get high precision: only predict one sample as positive where you are really sure
  - precision = 100% -> recall bad
- Calculate mean between recall and precision
  - Arithmetic mean not a good choice
    - precision = 100%, recall = 1% -> mean(precision, recall) = 50.5%
    - Too good a value for this bad classificator
  - Better: Harmonic mean - F1 value

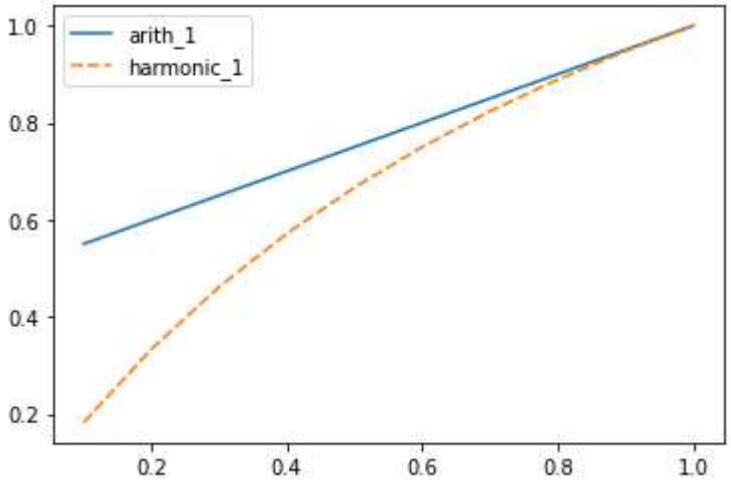
$$F1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Harmonic Mean vs Arithmetic Mean

Prof. Dr. Ingo Claßen

Sonntag, 5. Dezember 2021 18:07

	arith_1	harmonic_1
0.1	0.55	0.181818
0.2	0.60	0.333333
0.3	0.65	0.461538
0.4	0.70	0.571429
0.5	0.75	0.666667
0.6	0.80	0.750000
0.7	0.85	0.823529
0.8	0.90	0.888889
0.9	0.95	0.947368
1.0	1.00	1.000000

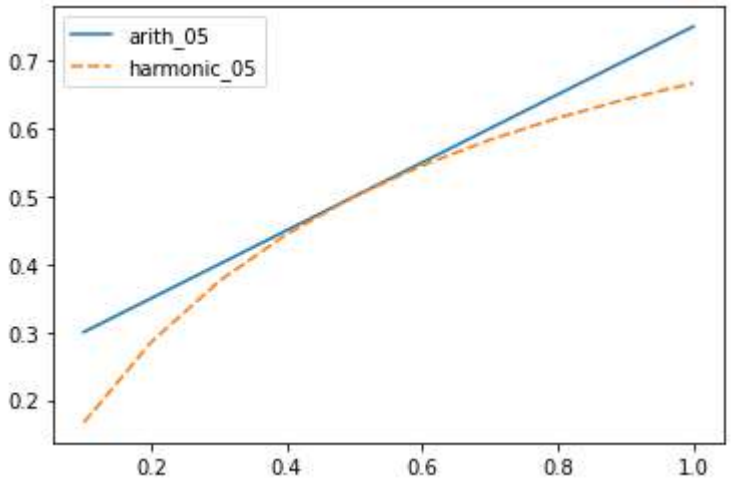


Recall

Precision = 1.0

---

	arith_05	harmonic_05
0.1	0.30	0.166667
0.2	0.35	0.285714
0.3	0.40	0.375000
0.4	0.45	0.444444
0.5	0.50	0.500000
0.6	0.55	0.545455
0.7	0.60	0.583333
0.8	0.65	0.615385
0.9	0.70	0.642857
1.0	0.75	0.666667



Recall

Precision = 0.5

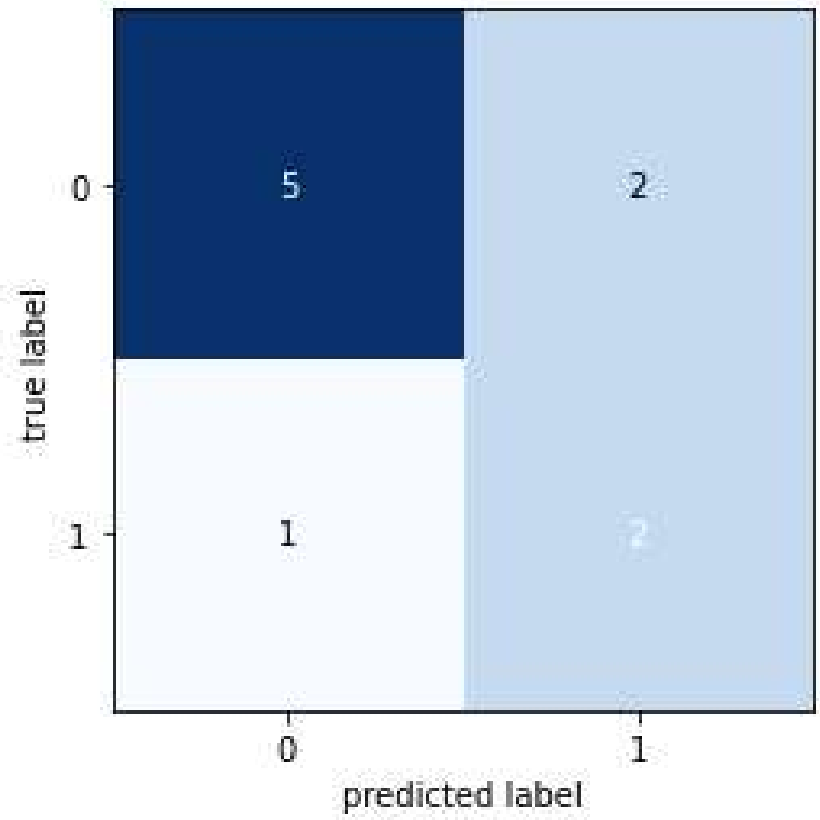
False Positive Rate -FPR (False Alarms)

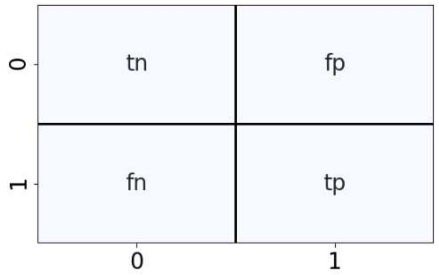
Prof. Dr. Ingo Claßen

Dienstag, 7. Dezember 2021 09:01

Fraction of false positive predicted to all negatives

true	1	1	1	0	0	0	0	0	0	0
pred	1	1	0	1	1	0	0	0	0	0



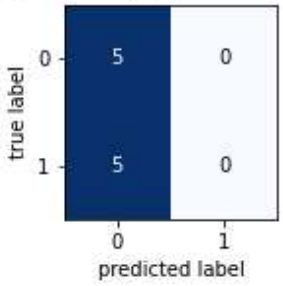


$$\text{fpr} = \frac{\text{fp}}{\text{fp} + \text{tn}}$$

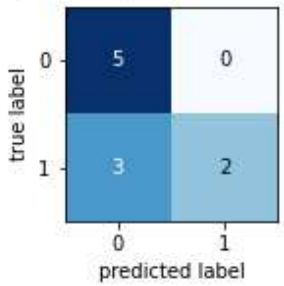
$$\text{fpr} = \frac{2}{2+5} = 0.29 = 29\%$$

true	1	1	0	1	1	0	0	0	1	0
score	0.9	0.8	0.7	0.6	0.55	0.4	0.39	0.38	0.37	0.1
pred >= 1.0	0	0	0	0	0	0	0	0	0	0
pred >= 0.8	1	1	0	0	0	0	0	0	0	0
pred >= 0.7	1	1	1	0	0	0	0	0	0	0
pred >= 0.5	1	1	1	1	1	0	0	0	0	0
pred >= 0.38	1	1	1	1	1	1	1	1	0	0
pred >= 0.37	1	1	1	1	1	1	1	1	1	0
pred >= 0.0	1	1	1	1	1	1	1	1	1	1

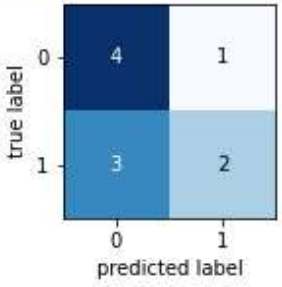
```
pred >= 1.0
true: [1 1 0 1 1 0 0 0 1 0]
pred: [0 0 0 0 0 0 0 0 0 0]
fpr: 0.0, tpr: 0.0
```



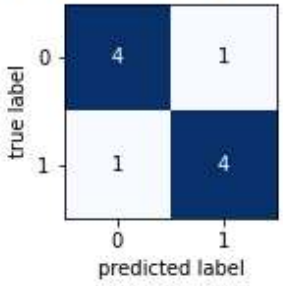

```
pred >= 0.8
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 0 0 0 0 0 0 0 0]
fpr: 0.0, tpr: 0.4
```



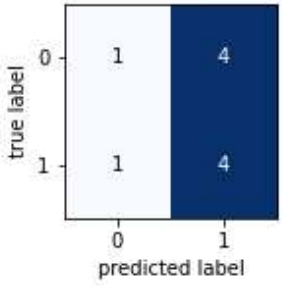

```
pred >= 0.7
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 1 0 0 0 0 0 0 0]
fpr: 0.2, tpr: 0.4
```



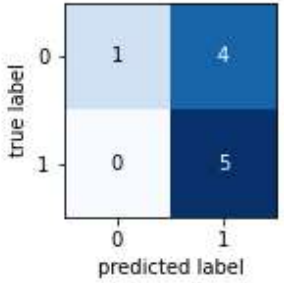

```
pred >= 0.5
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 1 1 1 0 0 0 0 0]
fpr: 0.2, tpr: 0.8
```



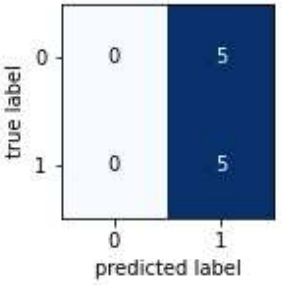

```
pred >= 0.38
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 1 1 1 1 1 1 0 0]
fpr: 0.8, tpr: 0.8
```

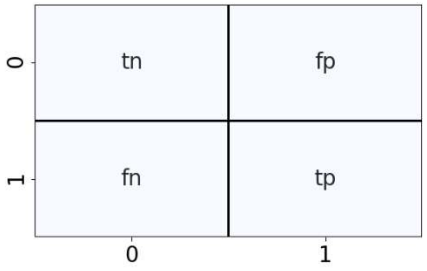



```
pred >= 0.37
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 1 1 1 1 1 1 1 0]
fpr: 0.8, tpr: 1.0
```

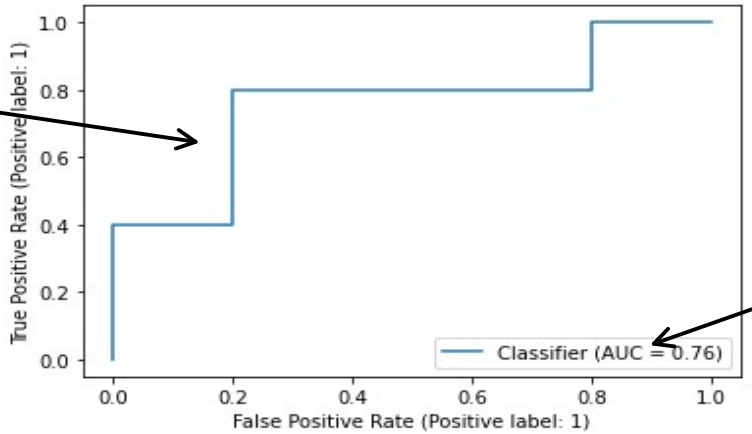



```
pred >= 0.0
true: [1 1 0 1 1 0 0 0 1 0]
pred: [1 1 1 1 1 1 1 1 1 1]
fpr: 1.0, tpr: 1.0
```





ROC Curve



Area under Curve

There is a threshhold that separates true and false

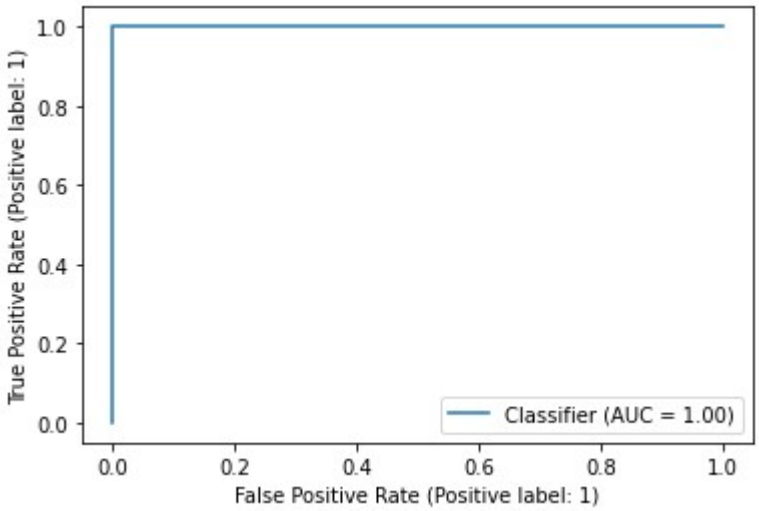
Example 1

true	1	1	1	1	1	0	0	0	0	0
score	0.9	0.8	0.7	0.6	0.55	0.4	0.39	0.38	0.37	0.1

Example 2

true	1	1	1	0	0	0	0	0	0	0
score	0.9	0.8	0.7	0.6	0.55	0.4	0.39	0.38	0.37	0.1

both have same ROC Curve



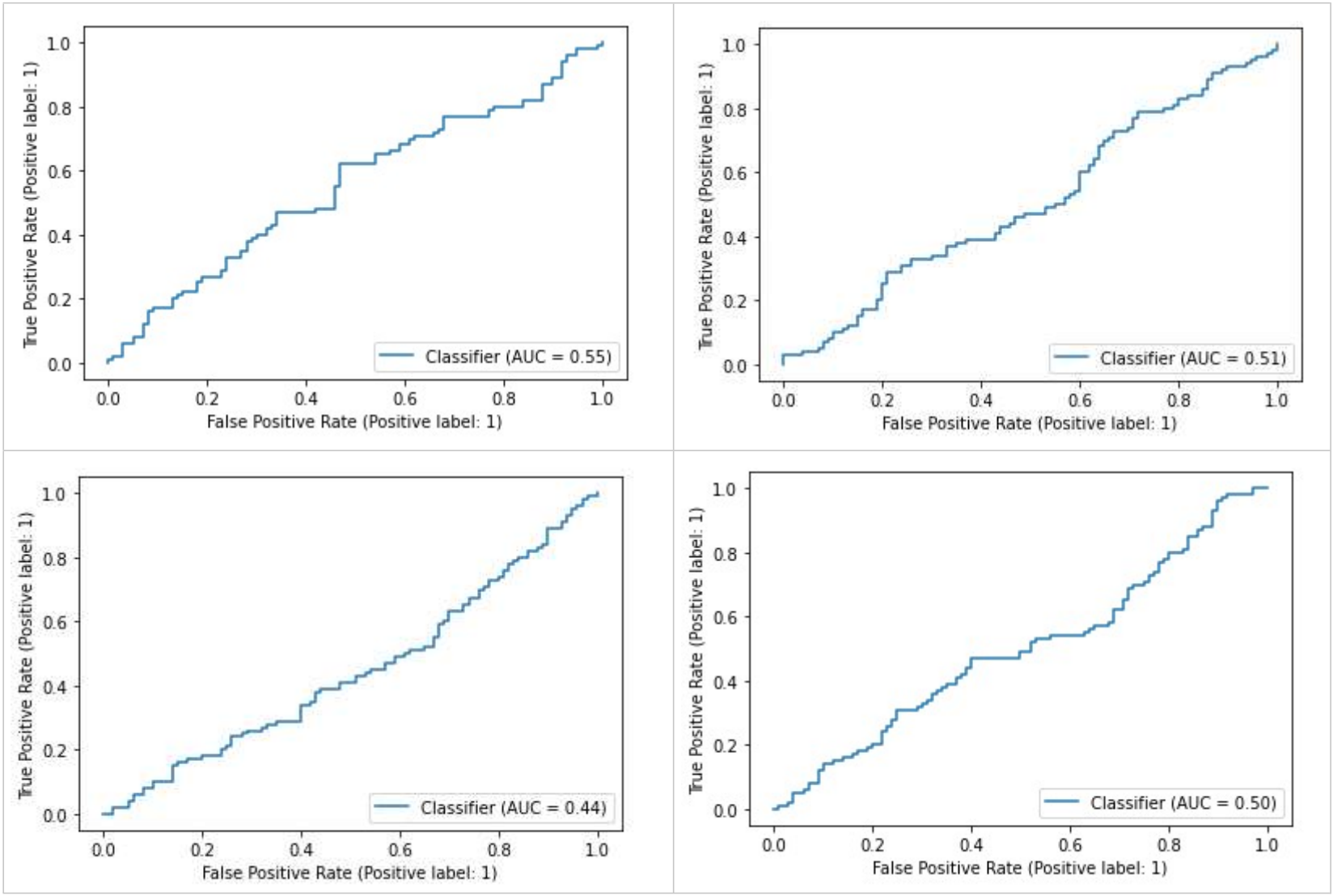
Random Classifier

Prof. Dr. Ingo Claßen

Donnerstag, 9. Dezember 2021 12:11

```
t7 = np.concatenate([np.ones(100), np.zeros(100)])
s7 = np.random.rand(200)
roc4 = RocCurveDisplay.from_predictions(t7, s7)
```

- Among all true, half of them are predicted positive and half of them negative
- Among all false, half of them are predicted positive and half of them negative



4 ROC Curves
(4 times generation of 200 random numbers)

AUC value near 0.5 means random classifier

Confusion Matrix (More than two Classes)

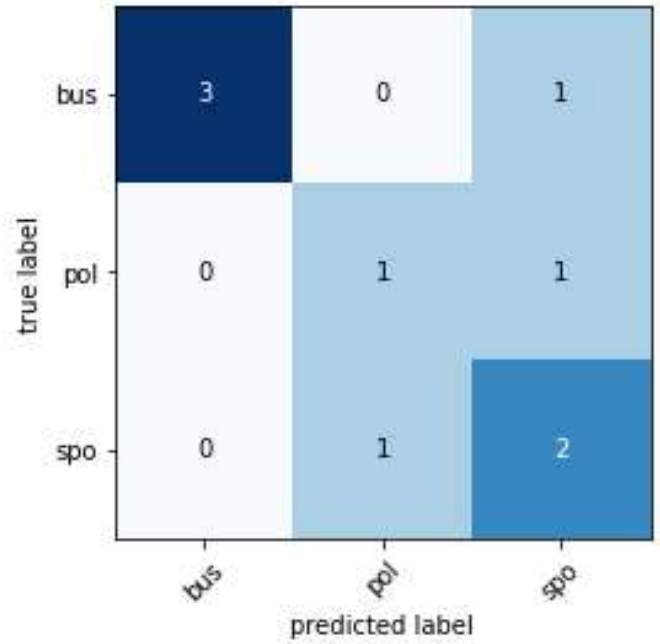
Prof. Dr. Ingo Claßen

Montag, 6. Dezember 2021 10:31

Classification of newspaper articles

bus	business
pol	politics
spo	sports

	1	2	3	4	5	6	7	8	9
true	pol	pol	spo	spo	spo	bus	bus	bus	bus
pred	pol	spo	spo	spo	pol	bus	bus	bus	spo



Accuracy

= all correctly classified / all

= sum(diagonal cells) / sum(all cells)

Precision / Recall / F1

per class

	precision	recall	f1	support
bus	1.0	0.750000	0.857143	4
pol	0.5	0.500000	0.500000	2
spo	0.5	0.666667	0.571429	3

averaged

	precision	recall	f1
macro	0.666667	0.638889	0.642857
micro	0.666667	0.666667	0.666667
weighted	0.722222	0.666667	0.682540

macro	average of values per class
micro	average per instances
weighted	weighted (wrt support) average of values per class