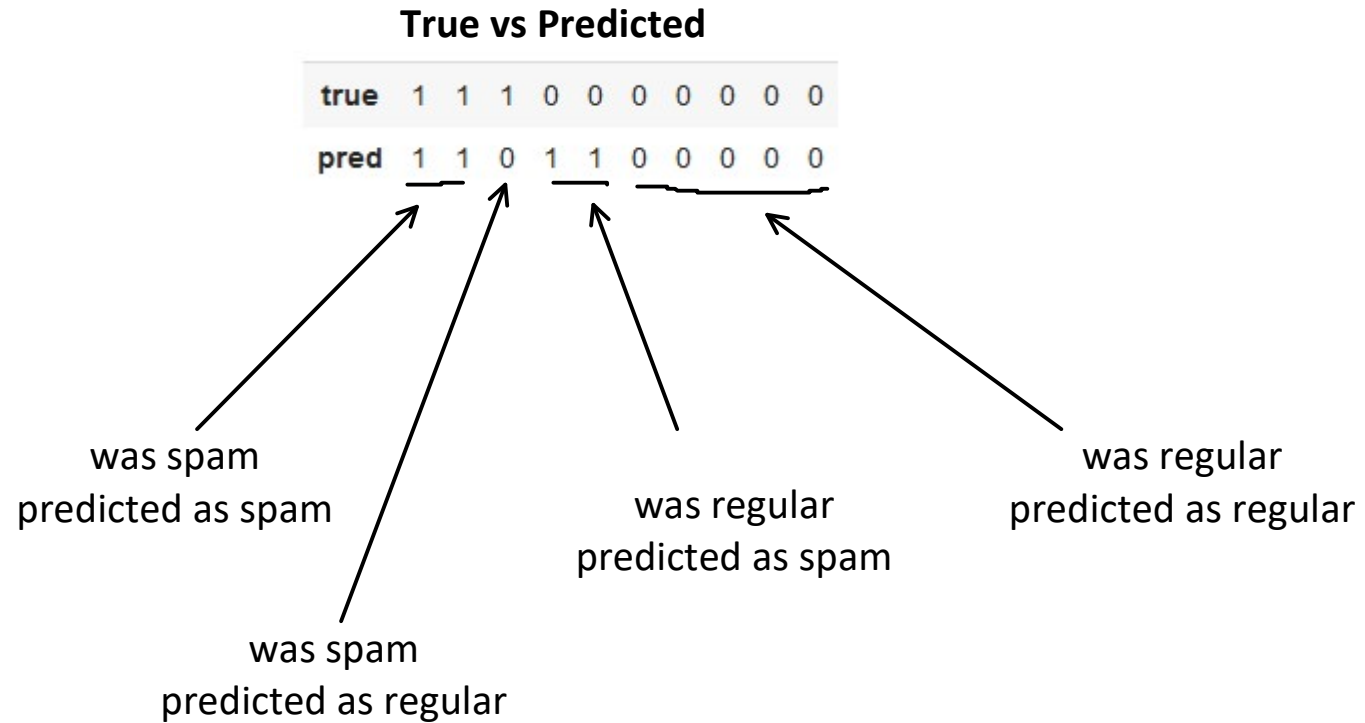


## Evaluation Metrics for Classification

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

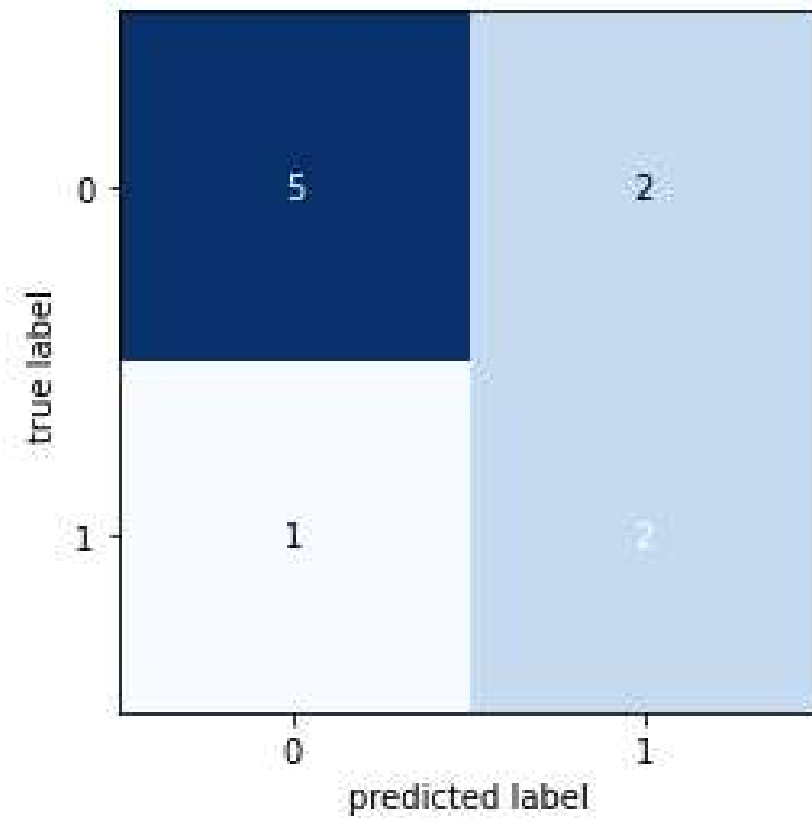
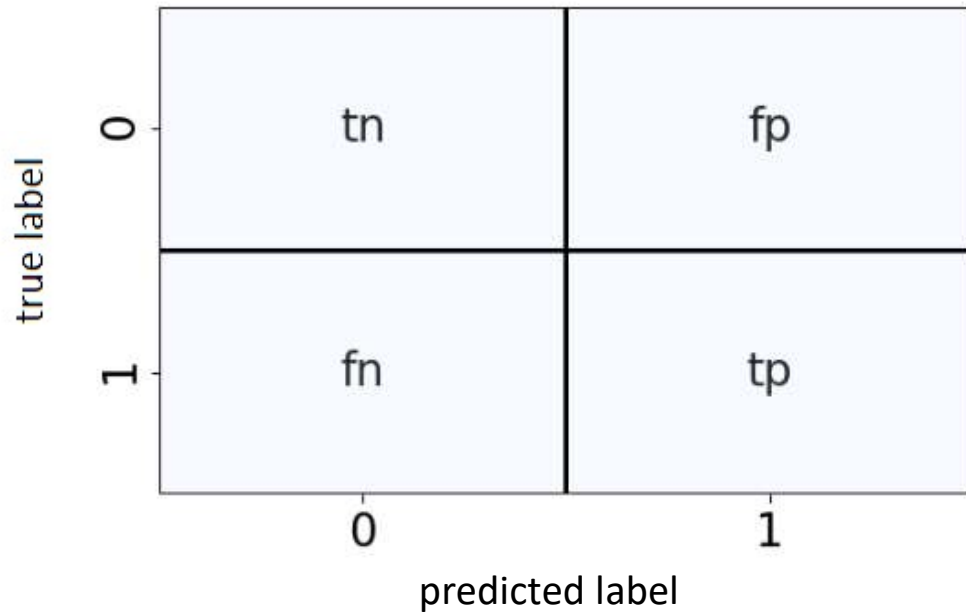


# Confusion Matrix (Two Classes)

Sonntag, 5. Dezember 2021 10:40

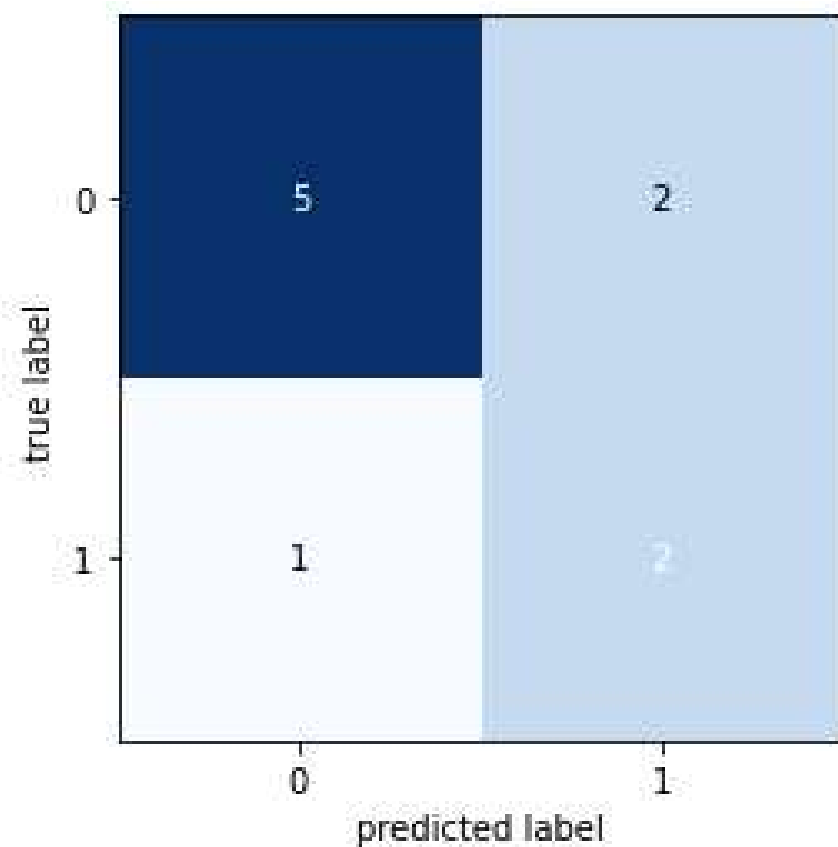
true label	predicted label	name	description
positive	positive	<b>tp</b>	true positive predicted
negative	negative	<b>tn</b>	true negative predicted
positive	negative	<b>fn</b>	false negative predicted
negative	positive	<b>fp</b>	false positive predicted

<b>true</b>	1	1	1	0	0	0	0	0	0
<b>pred</b>	1	1	0	1	1	0	0	0	0



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
```



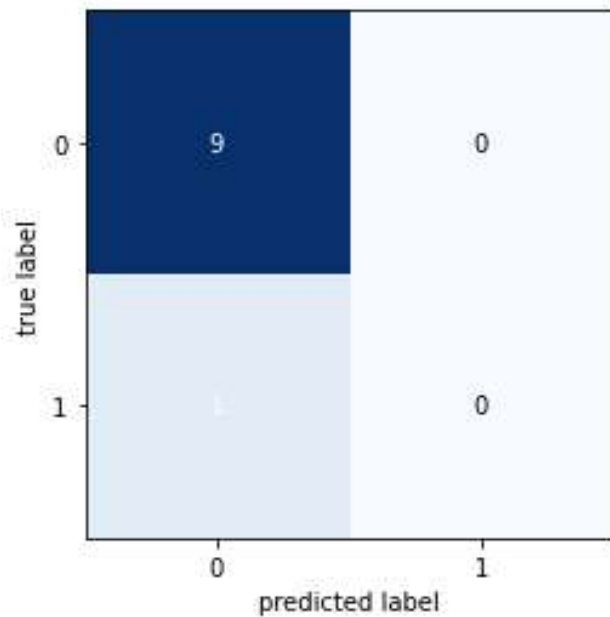
0	tn	fp
1	fn	tp
	0	1

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

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

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

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



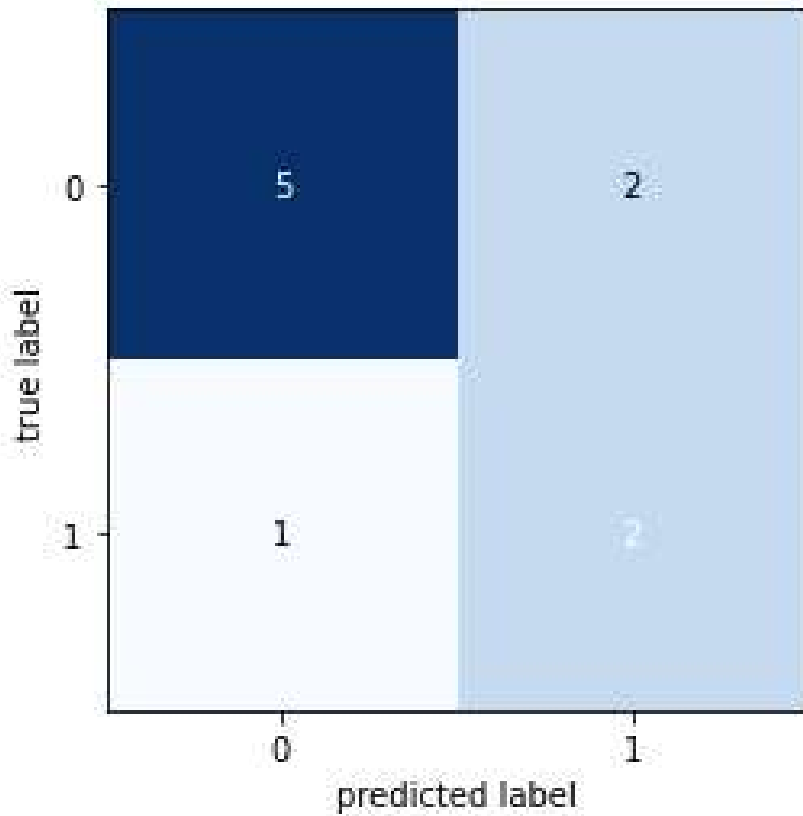
0	tn	fp
1	fn	tp
	0	1

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

but classifier is useless

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
```



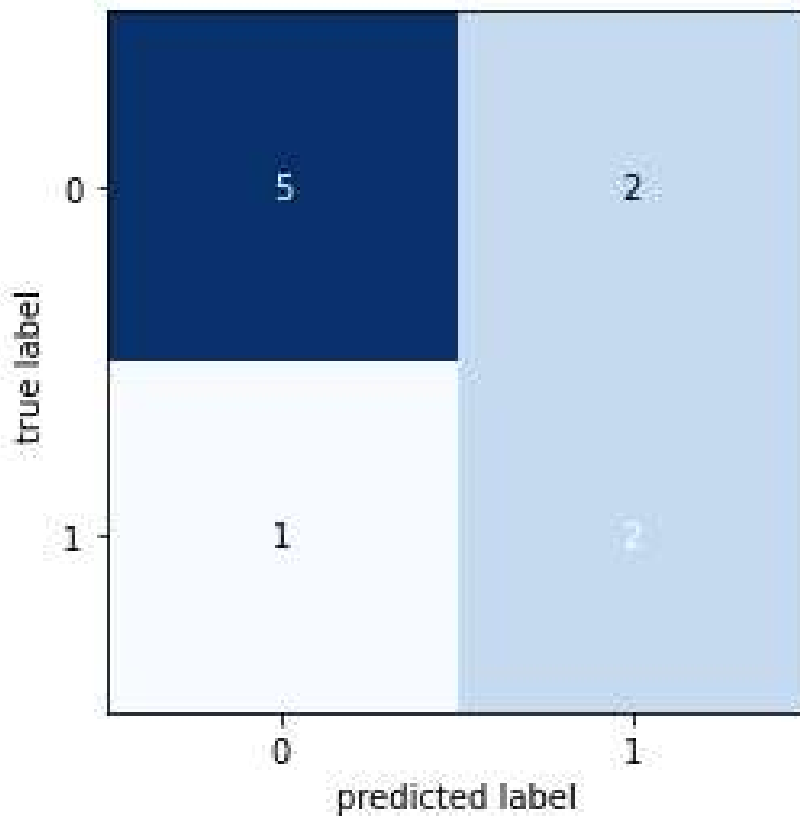
	0	1
0	tn	fp
1	fn	tp

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

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

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
```



0	tn	fp
1	fn	tp
	0	1

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

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

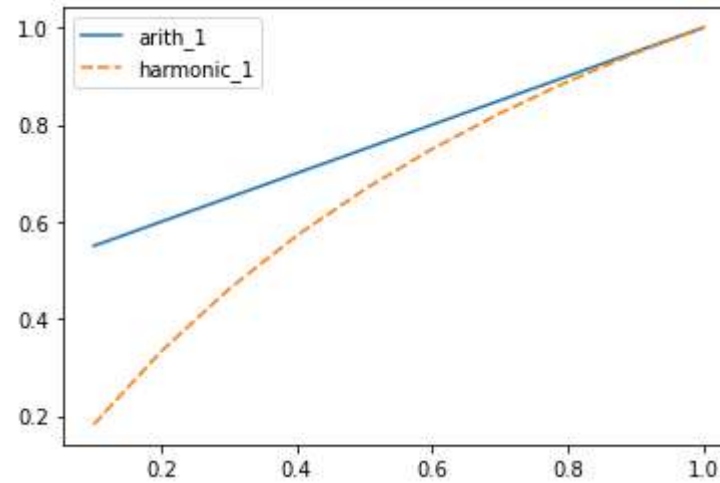
- Easy to get high recall: just predict positive
  - recall = 100% -> precision bad
- Easy to get high precision: only predict one sample as positive where you are really sure
  - precision = 100% -> recall bad
- Calculate mean between recall and precision
  - Arithmetic mean not a good choice
    -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

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



Sonntag, 5. Dezember 2021 18:07

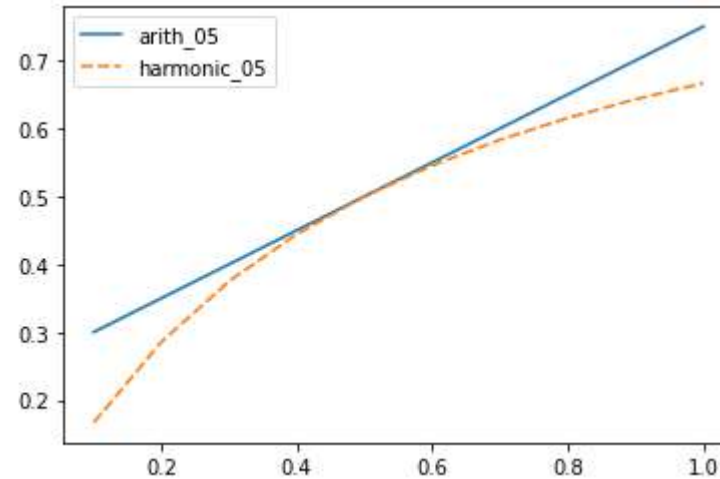
	arith_1	harmonic_1
<b>0.1</b>	0.55	0.181818
<b>0.2</b>	0.60	0.333333
<b>0.3</b>	0.65	0.461538
<b>0.4</b>	0.70	0.571429
<b>0.5</b>	0.75	0.666667
<b>0.6</b>	0.80	0.750000
<b>0.7</b>	0.85	0.823529
<b>0.8</b>	0.90	0.888889
<b>0.9</b>	0.95	0.947368
<b>1.0</b>	1.00	1.000000



Precision = 1.0

Recall

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

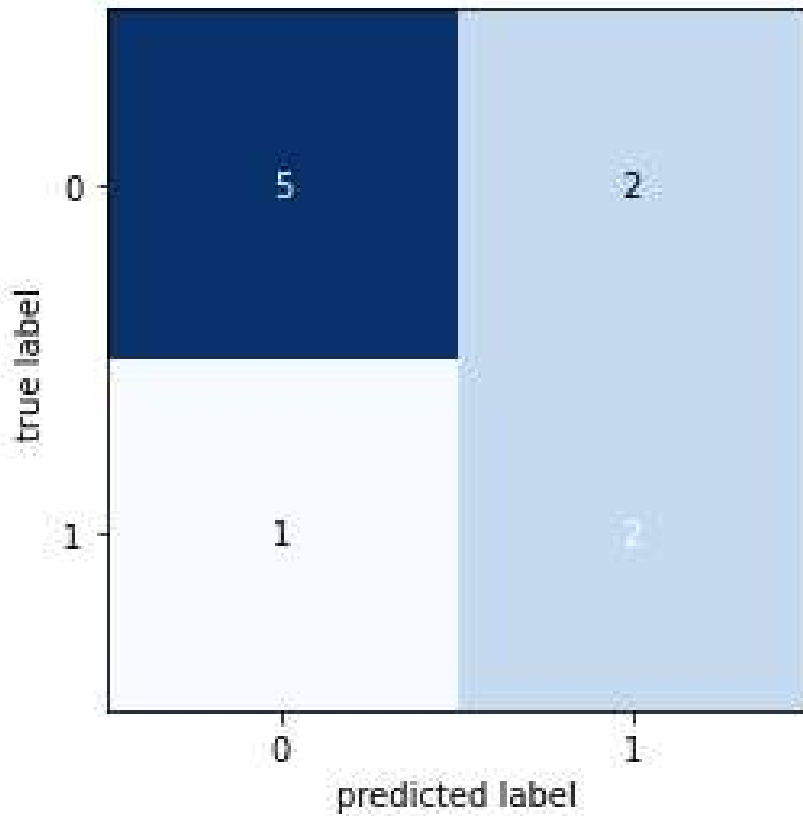


Precision = 0.5

Recall

Fraction of false positive predicted to all negatives

<b>true</b>	1	1	1	0	0	0	0	0	0
<b>pred</b>	1	1	0	1	1	0	0	0	0

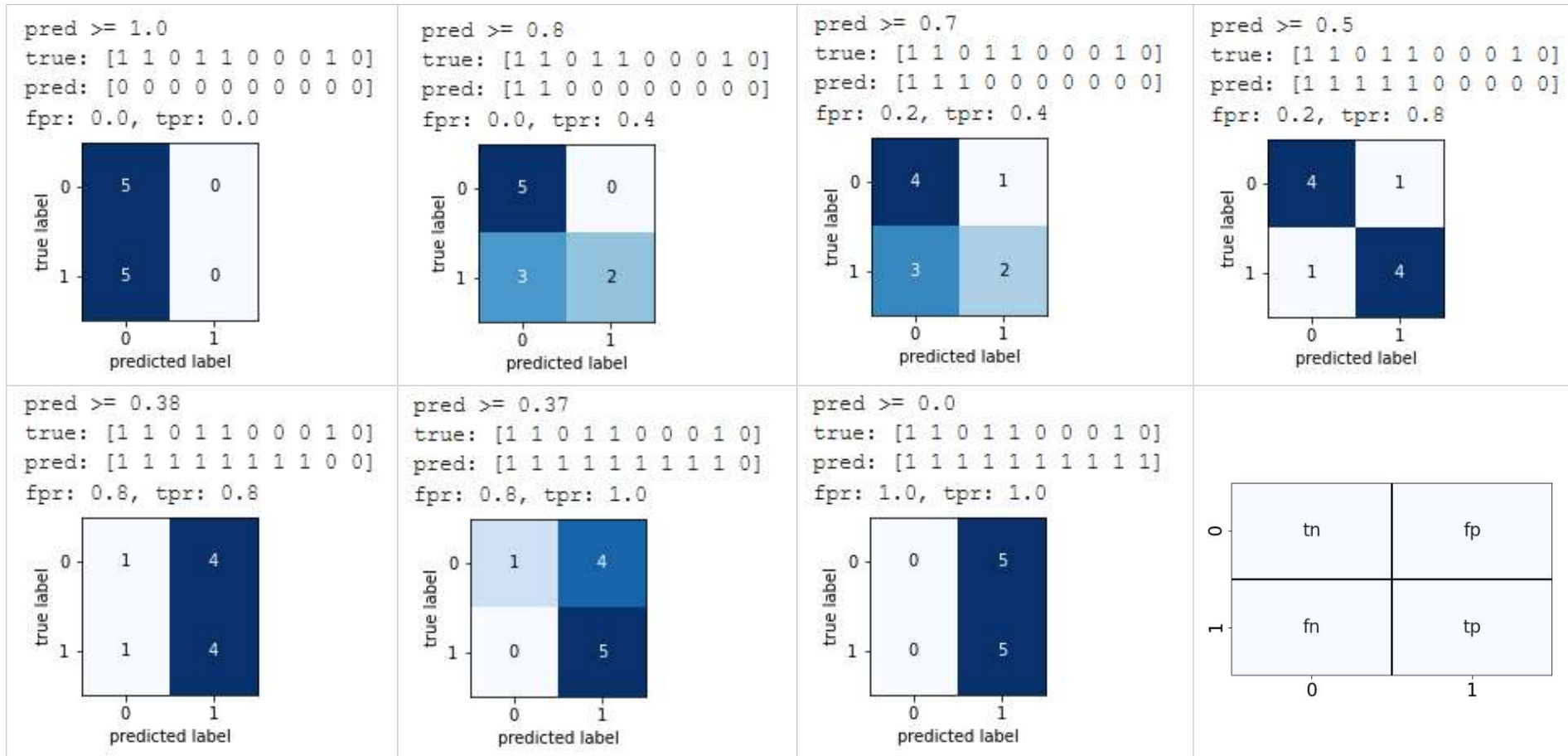


0	tn	fp
1	fn	tp
	0	1

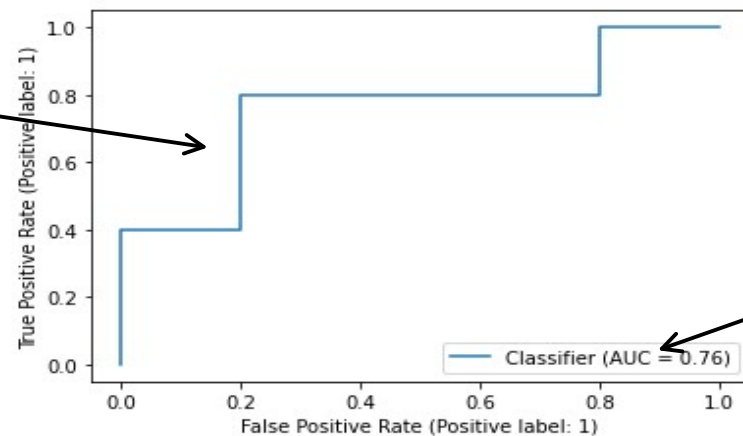
$$fpr = \frac{fp}{fp + tn}$$

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

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



ROC Curve



Area under Curve

There is a threshold that separates true and false

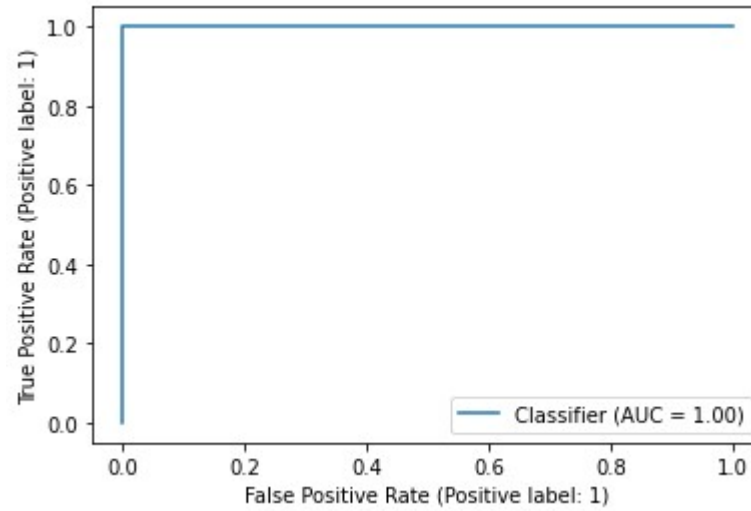
Example 1

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

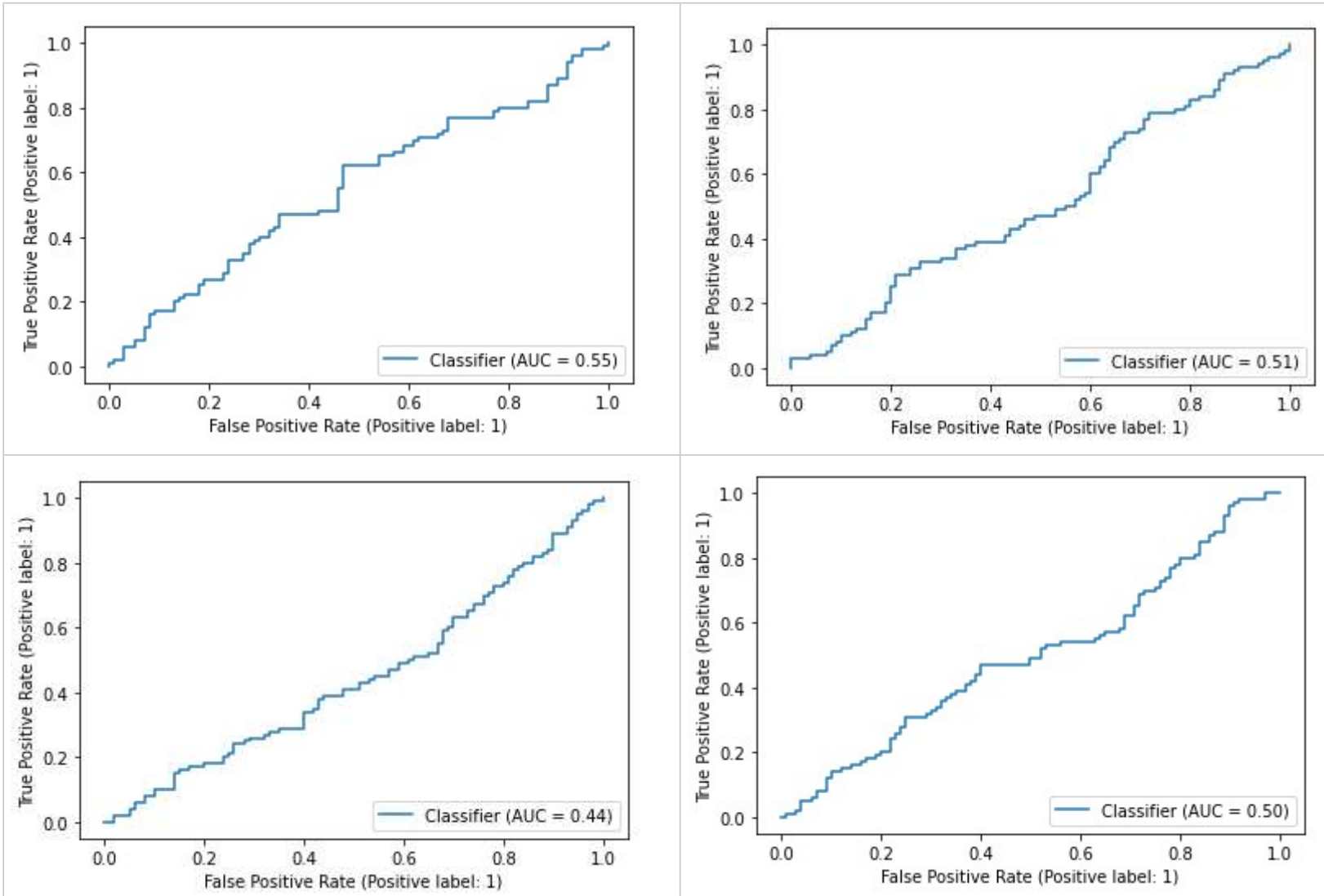
true	1	1	1	0	0	0	0	0	0	0
score	0.9	0.8	0.7	0.6	0.55	0.4	0.39	0.38	0.37	0.1

both have same  
ROC Curve



```
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



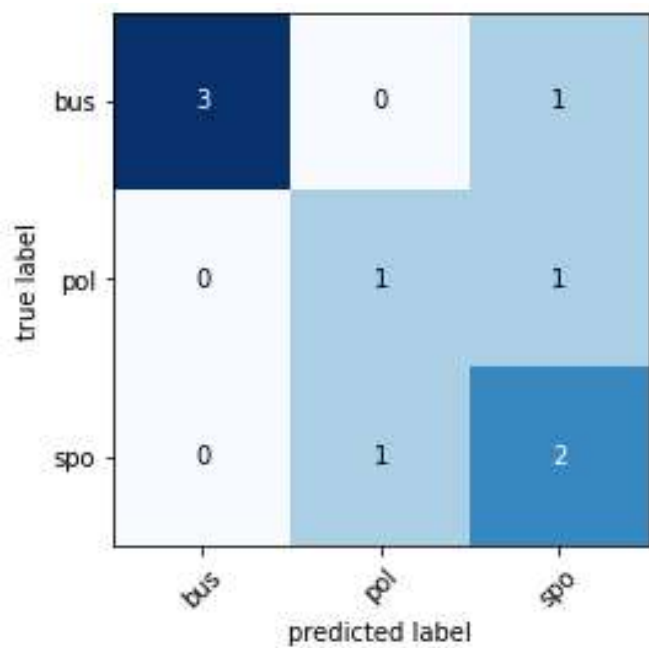
4 ROC Curves  
(4 times generation of 200 random numbers)

AUC value near 0.5 means random classifier

## Classification of newspaper articles

bus	business
pol	politics
spo	sports

	1	2	3	4	5	6	7	8	9
<b>true</b>	pol	pol	spo	spo	spo	bus	bus	bus	bus
<b>pred</b>	pol	spo	spo	spo	pol	bus	bus	bus	spo



## Accuracy

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

## Precision / Recall / F1

	precision	recall	f1	support
per class <b>bus</b>	1.0	0.750000	0.857143	4
<b>pol</b>	0.5	0.500000	0.500000	2
<b>spo</b>	0.5	0.666667	0.571429	3

	precision	recall	f1
averaged <b>macro</b>	0.666667	0.638889	0.642857
<b>micro</b>	0.666667	0.666667	0.666667
<b>weighted</b>	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