

Performance indices

Classification

Meunier Jean Christophe
November the 30th 2020, Bencode AI/Data science bootcamp

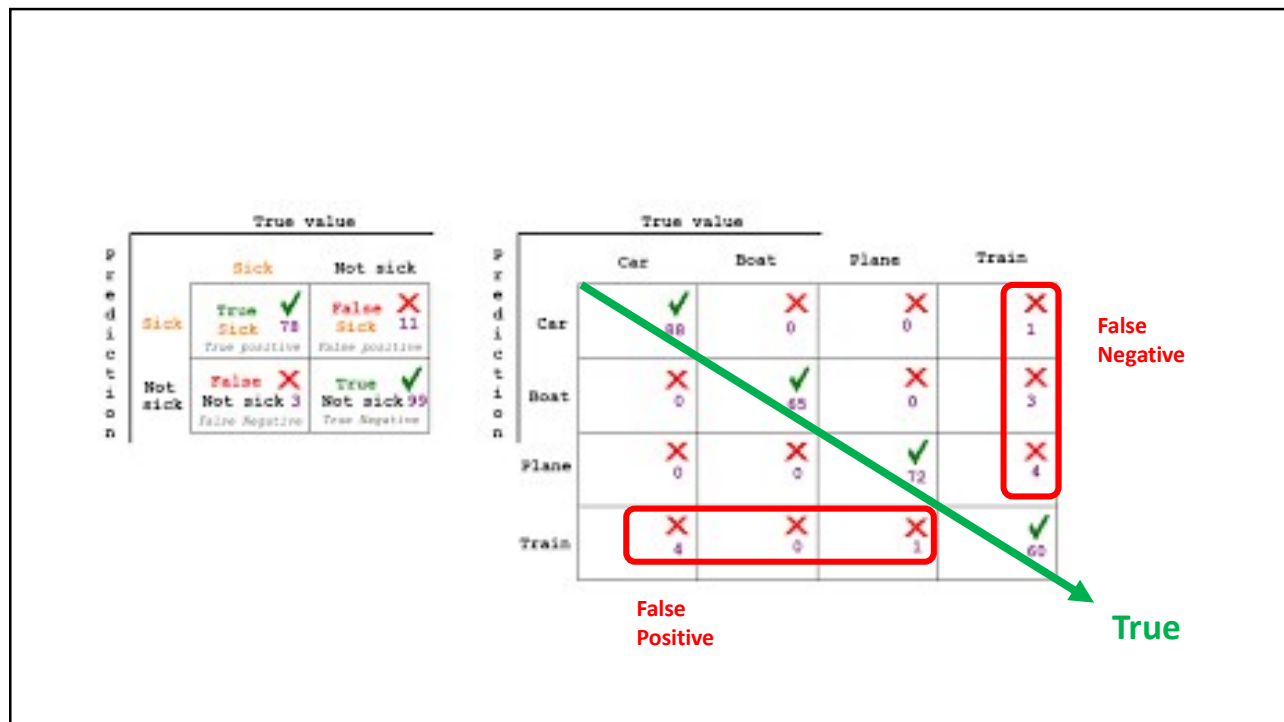
Classification

Confusion matrix

		Predicted Values	
		Class 1	Class 0
Actual Values	Class 1	True Positive	False Negative
	Class 0	False Positive	True Negative

Confusion matrix

		Predicted Values	
		Class 1	Class 0
Actual Values	Class 1	True Positive	False Negative
	Class 0	False Positive	True Negative



• Confusion !!!

		Predicted value					
		A	B				
Actual value	A	TP	FN	Predicted value	A	TP	FP
	B	FP	TN		B	FN	TN

- Scikit-learn
- Confusion matrix
 - Classification report

Accuracy

- Pourcentage de bonne prediction

		Predicted value		
		A	B	
Actual value	A	45	15	<p>Dans mon échantillon de 100 sujets</p> <ul style="list-style-type: none"> - 60 répondent positivement au vaccin - 40 répondent négativement
	B	13	27	

Bonnes prédictions

$$Accuracy = \frac{TP + TN}{Total} = \frac{45 + 27}{100} = 0,72$$

Accuracy !!! Si classes fortement asymétriques

- Modèle prédit que tout OK

		Predicted value		
		A	B	
Actual value	A	95	0	<p>Dans mon échantillon de 100 sujets</p> <ul style="list-style-type: none"> - 95 répondent positivement au vaccin - 5 répondent négativement
	B	5	0	

Bonnes prédictions

$$Accuracy = \frac{TP + TN}{Total} = \frac{95 + 0}{100} = 0,95$$

Accuracy !!! Si classes fortement asymétriques

- Modèle prédit que tout **KO**

		Predicted value		
		A	B	
Actual value	A	0	95	<p>Dans mon échantillon de 100 sujets</p> <ul style="list-style-type: none"> - 95 répondent positivement au vaccin - 5 répondent négativement
	B	0	5	

Bonnes prédictions

$$Accuracy = \frac{TP + TN}{Total} = \frac{0 + 5}{100} = 0,05$$

Précision

$$Precision = \frac{TP}{TP + FP} \rightarrow \text{Pénalise modèles trop optimistes}$$

- Ratio vrais positifs (TP) sur tous les positifs prédits (TP + FP)

		Predicted value		
		A	B	
Actual value	A	45	15	<p>Dans mon échantillon de 100 sujets</p> <ul style="list-style-type: none"> - 60 répondent positivement au vaccin - 40 répondent négativement
	B	13	27	

Bonnes prédictions

$$Precision = \frac{TP}{TP + FP} = \frac{45}{45 + 13} = 0,78$$

Recall

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

→ Pénalise modèles trop pessimistes

- Ratio vrais positifs (TP) sur tous les positifs 'réels' (TP + FN)

		Predicted value	
		A	B
Actual value	A	45	15
	B	13	27

Vaccin Covid

Dans mon échantillon de 100 sujets

- 60 répondent positivement au vaccin

- 40 répondent négativement

Bonnes prédictions

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{45}{45 + 15} = 0,75$$

Accuracy, precision, recall

Modèle pessimiste			
		Predicted values	
		A	B
Actual values	A	35	25
	B	2	38

Accuracy 0.73

Precision 0.95

La plupart des positifs du modèle sont des vrais positifs (peu de FP)

Recall 0.58
Beaucoup de réels positifs n'ont pas été correctement prédits (bcp de FN)

Modèle parfait			
		Predicted values	
		A	B
Actual values	A	60	0
	B	0	40

Accuracy 1.00

Precision 1.00

Recall 1.00

Modèle optimiste			
		Predicted values	
		A	B
Actual values	A	58	2
	B	25	15

Accuracy 0.73

Precision 0.69

Beaucoup de positifs du modèle ne sont pas des vrais positifs (bcp de FP)

Recall 0.97
La plupart des réels positifs ont été correctement prédits (peu de FN)

Réels positifs

Positifs du modèle

F1-score

- Moyenne harmonique de precision et recall
 - Moyenne harmonique
 - Inverse de la moyenne arithmétique des inverses des termes

$$\frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

- Quand on veut déterminer un rapport moyen (respectent liens de proportionnalité)
 - Ex : vitesse moyenne AR, quand vitesse A ≠ vitesse R

F1-score

- Moyenne harmonique de precision et recall

$$F1score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2}{\frac{precision + recall}{precision * recall}} = 2 * \frac{precision * recall}{precision + recall}$$

$$F1score = \frac{2}{\frac{1}{\frac{TP}{TP + FP}} + \frac{1}{\frac{TP}{TP + FN}}} = \frac{2}{\frac{TP + FP + TP + FN}{TP}} = \frac{2 * TP}{2 * TP + FP + FN} \rightarrow \text{\#FP et \#FN respecté dans le ratio/calcul de la Moyenne H}$$

Pénalise de manière équivalente modèles optimistes et pessimistes

F1-score

		Predicted value	
		A	B
Actual value	A	45	15
	B	13	27

$$Precision = \frac{TP}{TP + FP} = \frac{45}{45 + 13} = 0,78$$

$$Recall = \frac{TP}{TP + FN} = \frac{45}{45 + 15} = 0,75$$

$$F1score = 2 * \frac{precision + recall}{precision * recall} = 2 * \frac{0,78 * 0,75}{0,78 + 0,75} = 2 * \frac{0,585}{1,53} = 0,76$$

$$F1score = \frac{2 * TP}{2 * TP + FP + FN} = \frac{2 * 45}{2 * 45 + 13 + 15} = \frac{90}{118} = 0,76$$

Accuracy, Precision, Recall & F1-Score

Modèle pessimiste			
		Predicted values	
		A	B
Actual values	A	35	25
	B	2	38

Accuracy	0.73
Precision	0.95
Recall	0.58
F1-Score	0,72

Modèle parfait			
		Predicted values	
		A	B
Actual values	A	60	0
	B	0	40

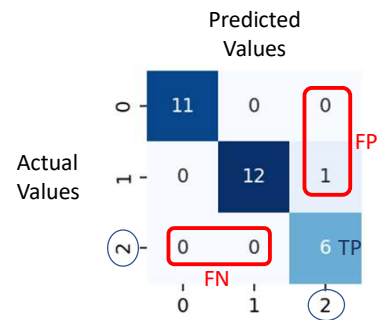
Accuracy	1.00
Precision	1.00
Recall	1.00
F1-Score	1.00

Modèle optimiste			
		Predicted values	
		A	B
Actual values	A	58	2
	B	25	15

Accuracy	0.73
Precision	0.69
Recall	0.97
F1-Score	0,81

Multiclass classification

```
from sklearn.metrics import confusion_matrix
conf_matrix = confusion_matrix(y_test, y_pred)
fig = plt.figure(figsize=(2,2))
sns.heatmap(conf_matrix, annot=True, cmap="Blues", cbar=False)
```



```
print(classification_report(y_test, y_pred, target_names=iris_data.target_names))
```

	precision	recall	f1-score	support
setosa	1.00	1.00	1.00	11
versicolor	1.00	0.92	0.96	13
virginica	0.86	1.00	0.92	6
accuracy	Moyenne simple			0.97
macro avg	0.95	0.97	0.96	30
weighted avg	0.97	0.97	0.97	30

Moyenne pondérée par taille des cellules