UNIVERSITY OF EDINBURGH
Business School

Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

Dr Xuefei Lu
The University of Edinburgh Business School

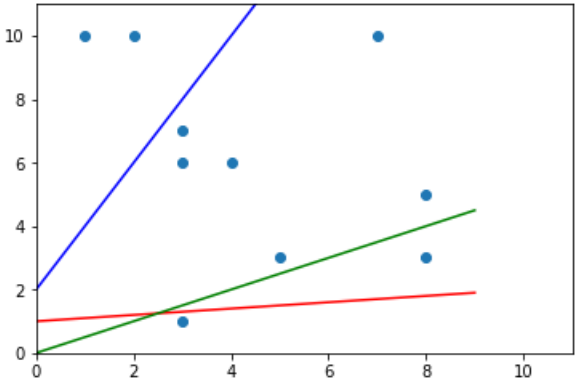
Evaluation Metrics

Evaluation Metrics

The Confusion Matrix

Evaluation metrics

Assume we are given a small dataset, like the blue points in the slide, which model (line) you will use for prediction?



Looking at error measures – scale dependent

Residuals: $e_i = y_i - \hat{y}_i$

Note that the error is still measured in the same dimension as the original value. That's why this measure cannot be used to compare errors, or even models that are measured in different dimensions.

- Mean error:

$$ME = \sum_{i=1}^{n} e_i$$

- Mean Absolute Error

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

- Root Mean Squared Error

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

- ME: negative and positive values cancel out
- RMSE: sensitive to outliers
- ME, MAE and RMSE are **scale-dependent**

Looking at error measures – scale independent

- The **percentage error**:

$$100\% \cdot \frac{e_i}{y_i}$$

is comparable for variables measured in different dimensions

- **Mean Absolute Percentage Error**:

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{e_i}{y_i}\right| \cdot 100\%$$

MAPE cannot be used when the actual value is 0

The minimum is always 0, while the upper bound is unlimited.

Looking at error measures – norminal values

- For nominal values used for cross-sectional/classification models,

$$e_i = \mathrm{I}_{\{\widehat{y_i} \neq y_i\}}$$

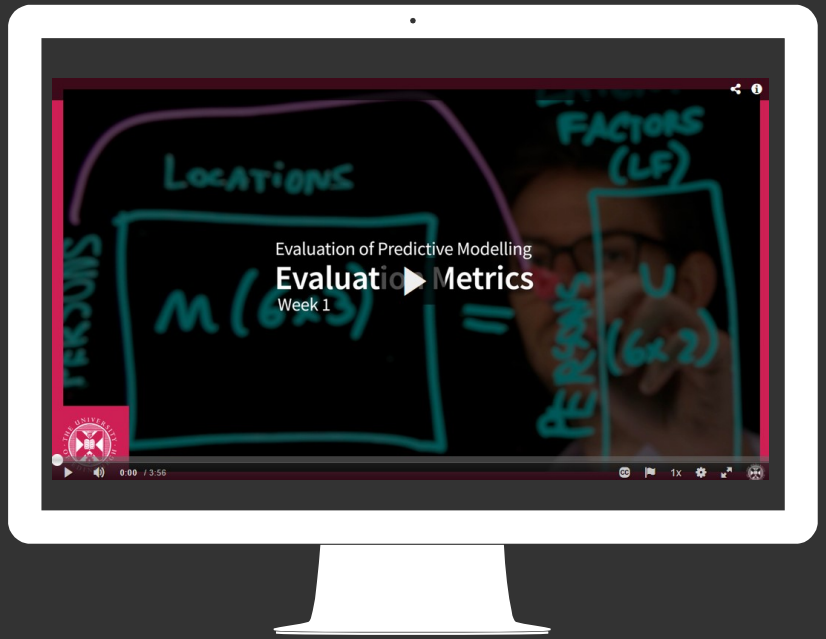
with $\mathrm{I}\{\cdot\}$ an indicator function returns value 1 for an incorrect prediction

- The Error Rate is then:

$$e = \frac{1}{n}\sum_{i}^{n} \mathrm{I}_{\{\widehat{y_i} \neq y_i\}}$$

the average number of incorrectly classified observations.

Evaluation Metrics



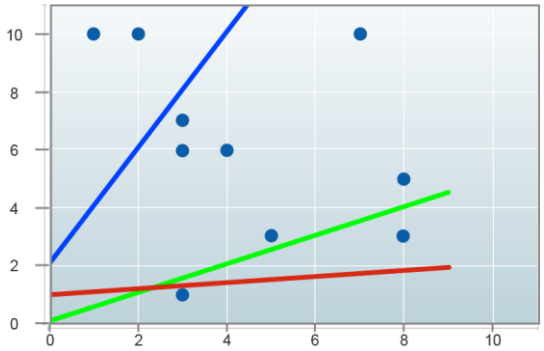
- Please watch the following video through this link

https://media.ed.ac.uk/media/Evaluation+Metrics/1_49sdvo92/114521421

Please complete the exercise in the following Jupyter file:
1 - Activity_1_calculating_error_measures.ipynb

To measure the performance of your simple linear regression models y = ax+b, calculate the ME, MAE, RMSE, and MAPE based on the distance between the predicted value and the actual value.

Activity: Calculating error measures

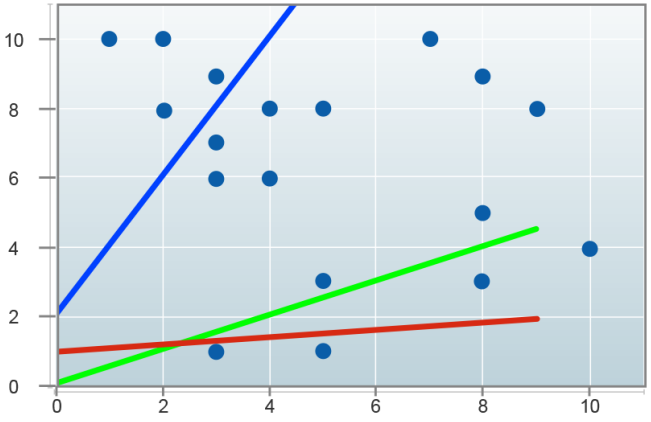


If define the following models
blue: y=2x+2 ,
green: y=0.1x+1
red: y=0.5x

	MAE	RMSE	MAPE (100%)
Blue	6.7	7.96	2.03
Red	4.72	5.62	0.68
Green	4.2	5.32	0.59

Activity: Calculating error measures – discussion

Now let's have a look at how the situation changes when we look at the extra data that is introduced.



	MAE	RMSE	MAPE (100%)
Blue	6.7	7.96	2.03
Red	4.72	5.62	0.68
Green	4.2	5.32	0.59
Blue (new)	7	8.72	2.66
Red (new)	5.02	5.72	70.18
Green (new)	4.3	4.92	70.76
Blue (All)	6.85	8.35	2.34
Red (All)	4.87	5.66	0.69
Green (All)	4.25	5.12	0.65

Activity: Calculating error measures – discussion

Evaluation Metrics

Evaluation Metrics

The Confusion Matrix

The Confusion Matrix – categorical/nominal

Assume we know how many times we have correctly predicted, and how many times we incorrectly predicted.

	Actual	Actual
Predicted	**True positive**	False positive
Predicted	False negative	**True negative**

	Actual = purchasing	Actual = not purchasing
Predicted = purchasing	**14**	2
Predicted = not purchasing	10	**5**

Is this a good prediction?

We have 19 (14+5) out of 31 (14+2+10+5), or 61.29% correct

BUT…

The Confusion Matrix – categorical/nominal

We can extend the confusion matrix for the multiclass problem: **the multiclass confusion matrix**

	Actual: class 1	Actual: class 2	…	Actual: class m
Pred.: class 1	c_{11}	c_{12}		c_{1m}
Pred.: class 2	c_{21}	c_{22}		c_{2m}
…				
Pred.: class m	c_{m1}	c_{m1}		c_{mm}

We can calculate the same metrics for class i as before:

$$TP_i = c_{ii}$$

$$FP_i = \sum_{j=1}^{m} c_{ij} - TP_i$$

$$FN_i = \sum_{j=1}^{m} c_{ji} - TP_i$$

$$TN_i = \sum_{j=1}^{m} \sum_{k=1}^{m} c_{jk} - TP_i - FP_i - FN_i$$

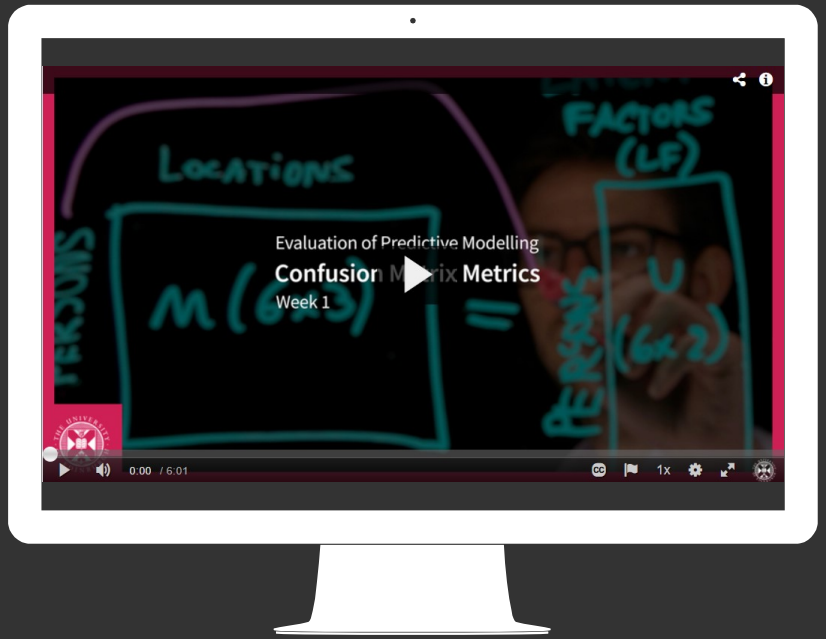
	Actual= i	Actual = not i
Predicted = i	c_{11}	$\sum_{j=1,j\neq i}^{m} c_{ij}$
Predicted = not i	$\sum_{j=1,j\neq i}^{m} c_{ji}$	$\sum_{j=1,j\neq i}^{m} \sum_{k=1,j\neq i}^{m} c_{jk}$

Confusion Matrix

- Please watch the following video through this link

https://media.ed.ac.uk/media/Confusion+Matrix/1_8k8ctgu2/114521421

Confusion Matrix Metrics



- Please watch the following video through this link

https://media.ed.ac.uk/media/Confusion+Matrix+Metrics/1_ngj4y5dv

Please try the following exercise
2 - Activity_3_code_your_own_confusion_matrix.ipynb

Activity: Code your own confusion matrix

Quiz

Based on the previous exercise:
2 - Activity_3_code_your_own_confusion_matrix.ipynb

Q1. The previous coding activity gave you an insight into the performance of the different classifications. What classification is performing best in terms of accuracy?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. They all performed similarly

Quiz

Based on the previous exercise:
2 - Activity_3_code_your_own_confusion_matrix.ipynb

Q2. What classification is performing best in terms of F1-score?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. They all performed similarly

Quiz

Based on the previous exercise:
2 - Activity_3_code_your_own_confusion_matrix.ipynb

Q3. What classification is performing best in terms of precision?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. Both classification #1 and #2 perform similarly

Please read the following Jupyter file
3 - Code_for_building_a_confusion_matrix.ipynb

Activity: Code for building a confusion matrix

Introducing to ROC and AUC curves

- Please watch the following video through this link

https://media.ed.ac.uk/media/ROC+Curve/1_bh76aiva/114521421

Please read the following Jupyter file
4 - Activity_4_ROC_curve_and_AUC.ipynb

Activity: ROC curve and AUC

Quiz

Q1. Consider the following metrics produced by applying classifiers to the three datasets we analysed:

Based on these metrics, which classifier do you think performed best?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. They all performed similarly

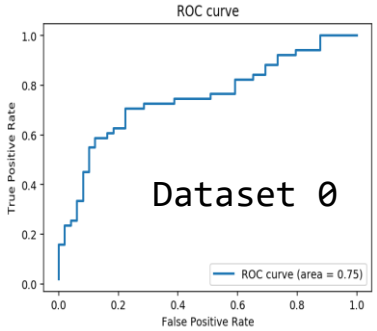
Dataset 0
Accuracy 0.68
Recall 0.7254901960784313
Precision 0.6727272727272727
Specificity 0.6326530612244898
Fall-out 0.3673469387755102
F1-score 0.6981132075471698

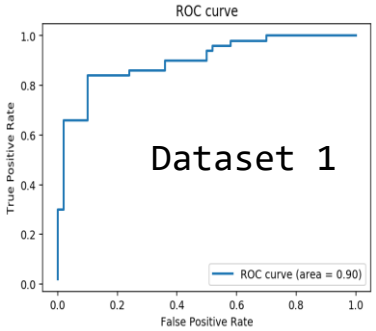
Dataset 1
Accuracy 0.86
Recall 0.84
Precision 0.875
Specificity 0.88
Fall-out 0.12
F1-score 0.8571428571428572

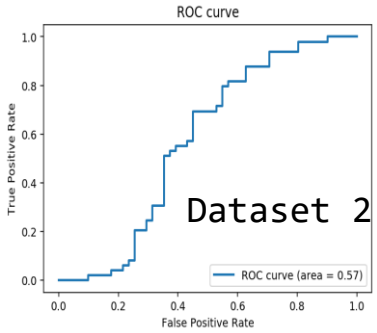
Dataset 2
Accuracy 0.91
Recall 0.9183673469387755
Precision 0.9
Specificity 0.9019607843137255
Fall-out 0.09803921568627451
F1-score 0.9090909090909091

Quiz

Q2. Consider these ROC curves of the three classifiers we used:







Based on these metrics, which classifier do you think performed best?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. They all performed similarly

Quiz

Q3. Considering the metrics used in the previous two questions, i.e., accuracy, F-score, and so on, and ROC/AUC, what classifier is performing the best overall?

- A. Classification #0 for Dataset 0
- B. Classification #1 for Dataset 1
- C. Classification #2 for Dataset 2
- D. They all performed similarly

UNIVERSITY OF EDINBURGH
Business School