Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

Dr Xuefei Lu
The University of Edinburgh Business School

Quiz

Q1. Which two of the following statements are true?

- A. Linear regression requires numeric inputs and continuous outputs.
- B. Linear regression requires numeric inputs and discrete outputs.
- C. Logistic regression requires numeric inputs and continuous outputs.
- D. Logistic regression requires numeric inputs and discrete outputs.

Quiz

Q1. Which two of the following statements are true?

- **A. Linear regression requires numeric inputs and continuous outputs.**
- B. Linear regression requires numeric inputs and discrete outputs.
- C. Logistic regression requires numeric inputs and continuous outputs.
- **D. Logistic regression requires numeric inputs and discrete outputs.**

Quiz

Q2. What makes a logistic regression a classification algorithm?

- A. Using probabilities
- B. Using a decision function
- C. Using a sigmoid function
- D. None of the above

Quiz

Q2. What makes a logistic regression a classification algorithm?

- A. Using probabilities
- **B. Using a decision function**
- **C. Using a sigmoid function**
- D. None of the above

Quiz

Q3. Which of the following statements is correct?

- A. The steeper the slope, the stronger the impact of the independent variable on the dependent

- B. The flatter the slope, the stronger the impact of the independent variable on the dependent

- C. The direction of the relationship is determined by the intercept.

- D. The distance between observations is measured by the intercept.

Quiz

Q3. Which of the following statements is correct?

- **A. The steeper the slope, the stronger the impact of the independent variable on the dependent**

- B. The flatter the slope, the stronger the impact of the independent variable on the dependent

- C. The direction of the relationship is determined by the intercept.

- D. The distance between observations is measured by the intercept.

Quiz

Q4. Why do we use the sigmoid function?

- A. Because it is the best way to model the log transform of odds

- B. Because it has a 0-1 outcome

- C. Because its shape forces outcomes to the edges of its range

- D. None of the above

Quiz

Q4. Why do we use the sigmoid function?

- A. Because it is the best way to model the log transform of odds
- **B. Because it has a 0-1 outcome**
- **C. Because its shape forces outcomes to the edges of its range**
- D. None of the above

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**

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
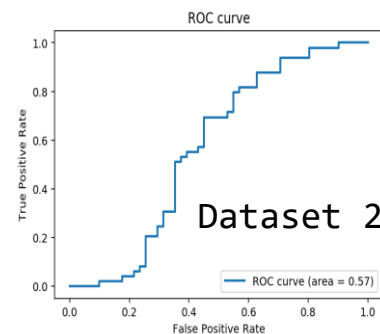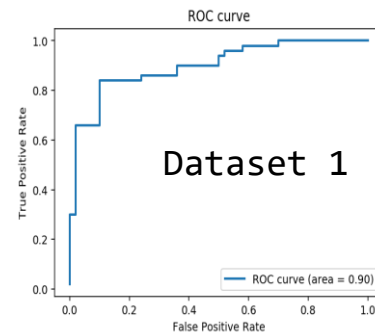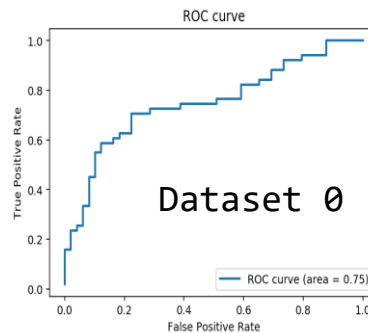Fall-out 0.09803921568627451
F1-score 0.9090909090909091

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