DATASCI 510

Data Science: Process and Tools

Lesson 10

Classification Accuracy

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Reflections

A scientist in his laboratory is not a mere technician: he is also a child confronting natural phenomena that impress him as though they were fairy tales.

Marie Skłodowska Curie, 1867-1934

Agenda

- Announcements: End of Class!
- Finish the Notebook for Lesson 9
- Fill the surveys
- Break
- An example with ROC and AuC
- Lesson_10_a_Classification_Accuracy.ipynb
- Break
- Lesson_10_b_Student.ipynb
- Interview question 1
- Interview question 2

Surveys

Please feel the survey!

Interview question 1 – Classification Accuracy

Actual Cases from Test Data:

- 20 got an intestinal infection in the hospital
- 20 did not get an intestinal infection in the hospital

Model Results (Predictions):

- 10 patients who were indeed infected each had a probability of **0.99**
- 10 patients who were indeed not infected each had a probability of **0.01**
- 10 patients who were not infected but each had a probability of **0.51**
- 10 patients who were infected but each had a probability of only **0.49**

What is the classification accuracy?

Interview question 1 – Classification Accuracy

Actual Cases from Test Data:

- 20 got an intestinal infection in the hospital
- 20 did not get an intestinal infection in the hospital

Model Results (Predictions):

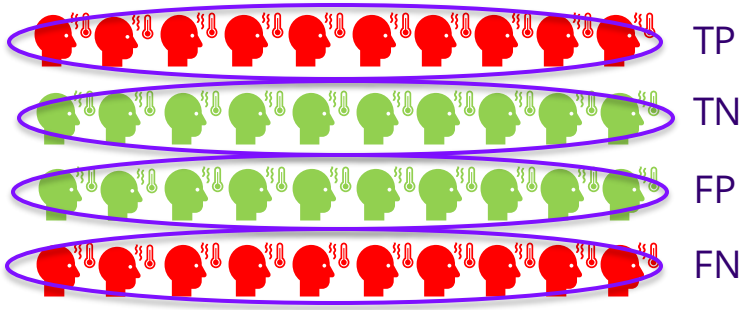
- 10 patients who were indeed infected each had a probability of **0.99**
- 10 patients who were indeed not infected each had a probability of **0.01**
- 10 patients who were not infected but each had a probability of **0.51**
- 10 patients who were infected but each had a probability of only **0.49**

Assuming threshold = 0.5

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{20}{40} = 0.5$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{10}{20} = 0.5$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{10}{20} = 0.5$$



Interview question 1 – Classification Accuracy

Actual Cases from Test Data:

- 20 got an intestinal infection in the hospital
- 20 did not get an intestinal infection in the hospital

Model Results (Predictions):

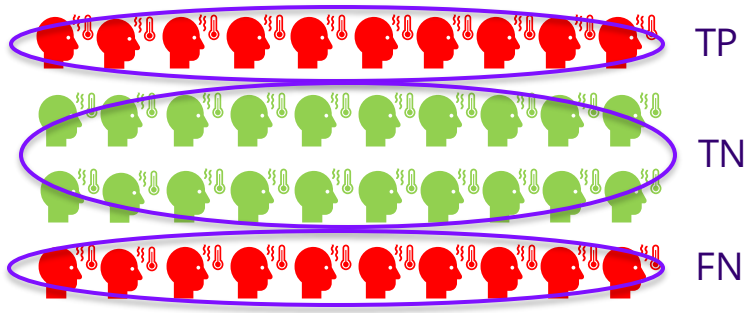
- 10 patients who were indeed infected each had a probability of **0.99**
- 10 patients who were indeed not infected each had a probability of **0.01**
- 10 patients who were not infected but each had a probability of **0.51**
- 10 patients who were infected but each had a probability of only **0.49**

Assuming threshold = 0.75

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{30}{40} = 0.75$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{10}{20} = 0.5$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{10}{10} = 1$$



Interview question 1 – Classification Accuracy

Actual Cases from Test Data:

- 20 got an intestinal infection in the hospital
- 20 did not get an intestinal infection in the hospital

Model Results (Predictions):

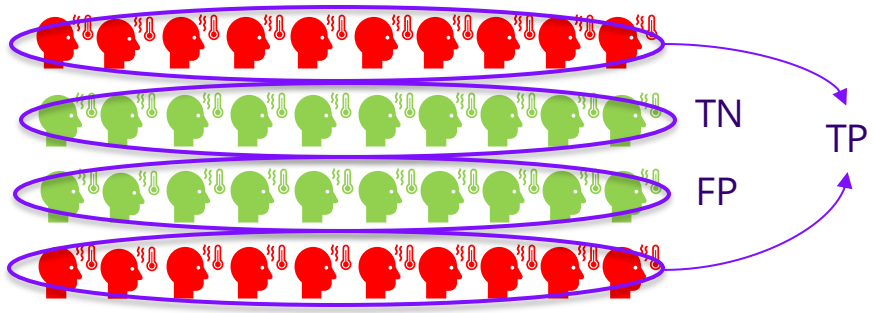
- 10 patients who were indeed infected each had a probability of **0.99**
- 10 patients who were indeed not infected each had a probability of **0.01**
- 10 patients who were not infected but each had a probability of **0.51**
- 10 patients who were infected but each had a probability of only **0.49**

Assuming threshold = 0.25

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{30}{40} = 0.75$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{20}{20} = 1$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{20}{30} = 0.67$$



Interview question 2 (from the last two weeks)

- Assume you are at one end of a tunnel, with a perfectly rectangular surface and need to move to the other end of the tunnel
- You are told that there are a given number of mines on the surface of this tunnel and their exact emplacements are known
- Furthermore, you know that it is safe to walk around the mines, as long as you keep a distance of R from the location of the mines
- Propose an algorithm to determine whether a safe path exists through the tunnel