

Nov 25, 2023

CSC461 – Assignment3 – Machine Learning

Ch. M. Kashan Akram

FA21-BSE-066

Use the dataset file “gender-prediction.csv” available on the shared Google Drive folder for this assignment.

PDF Report

Q1.

1. Number of instances: 110
2. Number of input attributes: 7
3. Number of possible values in the output attribute: 2
4. Number of categorical input attributes: 5
5. Class ratio (male vs female):
male 0.563636
female 0.436364

Q2.

How many instances are incorrectly classified?

- For the initial 2/3 train and 1/3 test split:
 - Logistic Regression: 13 instances incorrectly classified.
 - SVM: Around 16 instances incorrectly classified.
 - MLP: Around 11 instances incorrectly classified.
- After rerunning the experiment with an 80/20 train/test split:
 - Logistic Regression: Around 12 instances incorrectly classified.
 - SVM: Around 13 instances incorrectly classified.
 - MLP: Around 4 instances incorrectly classified.

When we rerun the experiment using a train/test split ratio of 80/20, do you expect to see any change in the results?

The accuracy has changed by approximately 3.5%.

The number of incorrectly classified instances may vary by approximately 9 instances.

Why are beard and scarf the most “powerful” in the prediction task?

These both are vital in determining gender due to societal rules. The exclusion of 'beard' and 'scarf' would result in roughly 8% decrease in accuracy for the models.

When we try to exclude these 2 attributes from the dataset and rerun the experiment (using 80/20 train/test split), do you expect to find any change in the results?

Yes, there might be a change. There's a 6% decrease in accuracy for the models when 'beard' and 'scarf' are excluded.

Q3.

P=5

Monte Carlo Cross-Validation F1 Scores: The F1 scores for each fold are around 1%. Leave

P-Out Cross-Validation F1 Scores: Around 3.4%.

Q4.

72 155 no medium 40 no green female

65 140 yes short 39 yes blue female

68 180 no long 42 yes brown male

69 160 no medium 38 yes gray female

70 145 yes bald 40 no black male

63 130 no short 37 no brown female

71 195 yes medium 43 yes gray male

66 150 no medium 38 no blue male

64 135 no long 37 yes green female

67 170 yes medium 42 no black male

Accuracy: The accuracy has changed by approximately 1.2%. Precision: Precision is fluctuated by approximately 2%. Recall: Recall is approximately 1.5%.