REPORT MACHINE LEARNING AND PATTERN RECOGNITION

Fingerprint spoofing detection

Alessia Intini

s309895

Politecnico di Torino

2023/24

Contents

The goal of project is to perform a binary classification on fingerprint spoofing detection, that is to identify genuine vs counterfeit fingerprint images. The dataset consists of labeled samples corresponding to the genuine (True, label 1) class and the fake (False, label 0) class, the data is 6-dimensional.

1 Dataset Analysis

1.1 Training and evaluation sets

The datasets provided contain 6000 samples; the first 6 values in each row represent the features, while the last value is the label. Specifically they are:

- Training Set: 2990 samples beloging to the Fake class (label 0) and 3010 samples belonging to the Genuine class (label 1)

- Evaluation Set: 3010 samples beloging to the Fake class (label 0) and 2990 samples belonging to the Genuine class (label 1)

We will use the Training Set to perform all the analysis, during this phase the dataset is divided to use about 60% of it as the training dataset and the remaining 40% for validation. After this step, the most promising models were chosen and the evaluation dataset was used to evaluate them and make the final considerations.

1.2 Features analysis

- Feature 1 and feature 2
  Yes the class overlap feature 1 in range x [-2.665,2.237], y[0,0.315] and for feature 2 in range x [-2.670,3.000] and y[0,0.325].
  The mean are similar in two feature, one is 0.00170711 and for feature two is 0.00503903.
  Yes the variance is similar 1.00134304, 0.9983527.
  Modes of hist in feature 1 for false value is y 0.541 for x in [-0.213,0.276], while for feature 2 is also for false value y 0.516, x [-0.402,0.165]

- Feature 3 and feature 4
  Yes the class overlap feature 3 in range x [-2.009,1.190], y[0,0.309] and for feature 4 in range x [-1.684,1.816] and y[0,0.371].
  The mean are similar in two feature, one is -0.00560753 and for feature two is 0.00109537.
  Yes the variance is similar 1.0024818 , 0.99029389.
  Modes of hist in feature 3 for false value is y 0.517 for x in [-1.063,-0.568], while for feature 4 is also for false value y 0.525, x [0.290,0.783]

- Feature 5 and feature 6
  Yes the class overlap feature 5 in range x [-2.066,2.004], y[0,0.282] and for feature 6 in range x [-2.000,2.180] and y[0,0.301].
  The mean are similar in two feature, one is -0.00700025 and for feature two is 0.00910515.
  Yes the variance is similar 1.00119747 , 0.99722374.

Modes of hist in feature 5 for true value is y 0.572 for x in [-1.211,-0.783], while for feature also for true value 6 y 0.553, x [-1.273,-0.817]

2   Dimensionality reduction

2.1   PCA

2.2   LDA

3   Multivariate Gaussian Density

4   Classification model analysis

4.1   Gaussian Models

4.1.1   MVG Gaussian Classifier

4.1.2   Naive Bayes Classifier

4.1.3   Tied Bayes Classifier

4.1.4   Gaussian Models Comparison

4.2   Logistic Regression Classifier

4.2.1   Quadratic Logistic Regression(QLR)

4.3   SVM Classifier

4.3.1   Linear SVM

4.3.2   Kernel SVM

4.4   GMM Classifier

5   Score Calibration

5.1   Calibration Analysis on Selected Models

5.2   Calibrating Scores for Selected Models

6   Experimental Results

6.1   Calibration on evaluation score

6.2   Considerations

7   Conclusions