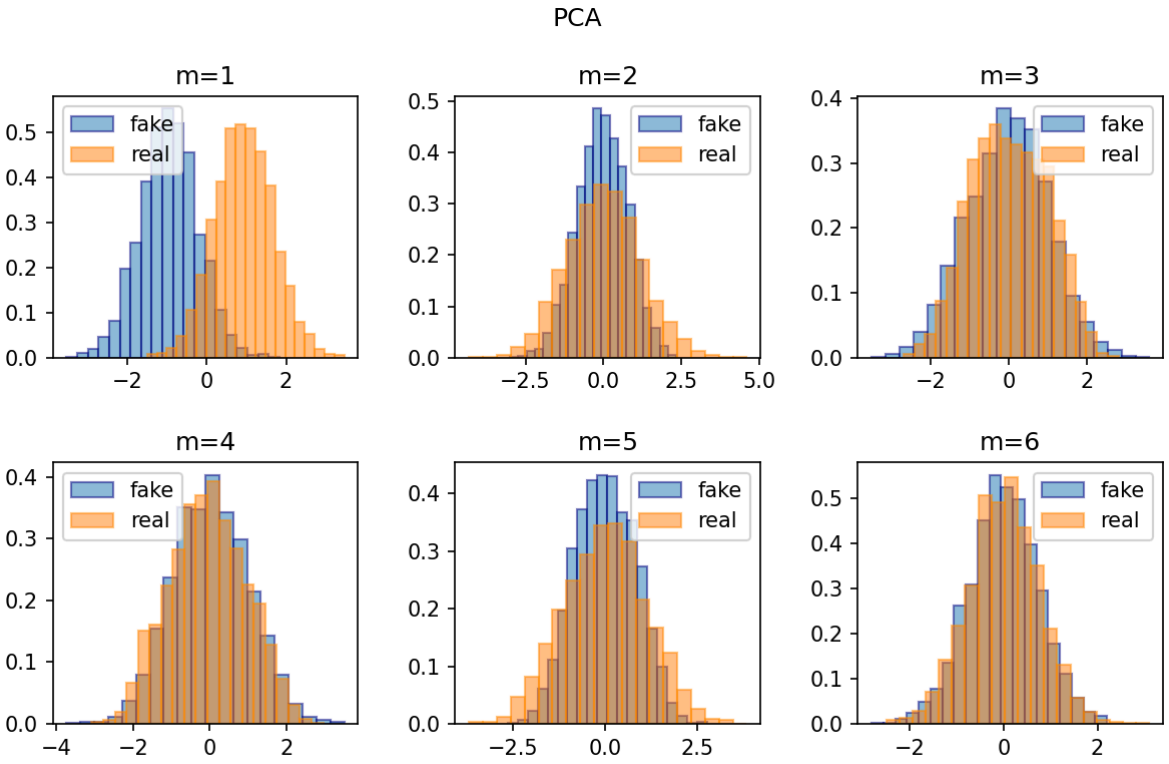
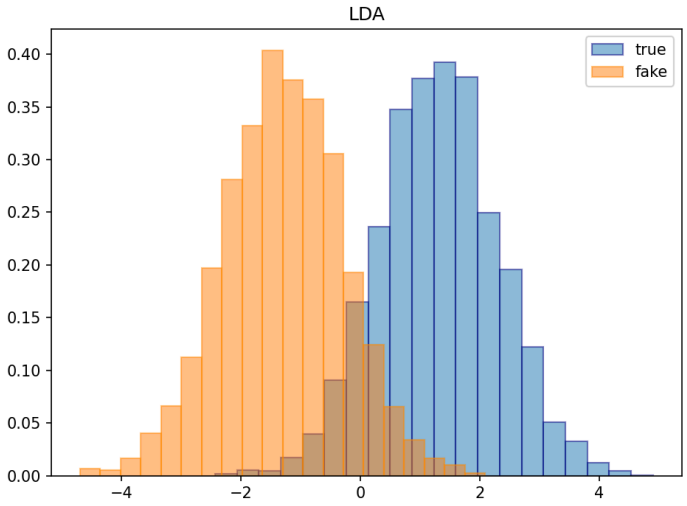
LAB03 REPORT

applying PCA can help avoiding the curse of dimensionality in dataset with a lot of features. In our dataset we have 6 features so we can choose m=1,2,3,4,5,6.



As we can see, along PCA principal direction (m=1), the two classes are very distinct and easy to notice, and grow the directions, they start to collide with each other.

Now we analyze the 1 dimensional LDA



LDA also does a good job at separating the two classes, it is very similar as PCA with m=1, there is very little overlap compared to the previous lab.

We now try to use LDA as a classifier, this is doable but we have to choose a threshold (greater than it: class 1, lower than it: class 0). we do so by computing the average of the projected class mean and this is the result.

```
threshold:0.01853437678620717 4
----------LDA without preprocessing----------
Number of erros: 186 (out of 2000 samples)
Error rate: 9.3%
```

An error rate of 9.3% is not bad. Now let's try to change change the threshold to 1

```
threshold:1
----------LDA without preprocessing----------
Number of erros: 387 (out of 2000 samples)
Error rate: 19.4%
```

The error rate is greater, this is expected because a bigger threshold will favor one class over the other.

```
threshold:0.11
----------LDA without preprocessing----------
Number of erros: 183 (out of 2000 samples)
Error rate: 9.2%
```

With some trial and error I found this threshold which gives me a little bit less error, it doesn't improve the performance much so this means that the average of the projected class mean gives us already a good discriminant threshold for the two classes.

For the final part of the project we had to pre-process the dataset with PCA before applying LDA for the classification. This can be very helpful when there is a lot of dimensionality but in our case the features space had only 6 dimensions so I don't know how much it will help to improve it. Let's try anyway:

```
----------LDA with PCA pre-processing, using m=1----------
Number of erros: 187 (out of 2000 samples)
Error rate: 9.3%
----------LDA with PCA pre-processing, using m=2----------
Number of erros: 185 (out of 2000 samples)
Error rate: 9.2%
----------LDA with PCA pre-processing, using m=3----------
Number of erros: 185 (out of 2000 samples)
Error rate: 9.2%
----------LDA with PCA pre-processing, using m=4----------
Number of erros: 185 (out of 2000 samples)
Error rate: 9.2%
----------LDA with PCA pre-processing, using m=5----------
Number of erros: 186 (out of 2000 samples)
Error rate: 9.3%
```

As we expected, PCA pre-processing didn’t help much, it reduces of 1-2 errors, with the best m being 2/3/4.