Yifan Yao
ECE 532
Gender classification
Email: yyao66@wisc.edu

Project Report for Gender Classification by Facial Feature

Introduction:

When meet a new person, most of time people determine one's gender by looking at face. Some of them determine it by single feature, like compare with male, more female has long hair. Some of them determine it by combine with more than one feature. Thus, the purpose for this project is determine people's gender by using facial features.

Project dataset:

The dataset for this project is from Kaggle (https://www.kaggle.com/elakiricoder/gender-classification-dataset). It is labeled data, which has 7 features like long of hair, width and height of forehead, long and wide of nose, lips thin and distance from nose to lips for 5001 samples. For easy to classification, set the label male to 1 and female to -1. The dataset shows below:

	long_hair	forehead_width_cm	forehead_height_cm	nose_wide	nose_long	lips_thin	distance_nose_to_lip_long	gender
0	1	11.8	6.1	1	0	1	1	1
1	0	14.0	5.4	0	0	1	0	-1
2	0	11.8	6.3	1	1	1	1	1
3	0	14.4	6.1	0	1	1	1	1
4	1	13.5	5.9	0	0	0	0	-1
...	...	...	...	...	...	...	...	...
4996	1	13.6	5.1	0	0	0	0	-1
4997	1	11.9	5.4	0	0	0	0	-1
4998	1	12.9	5.7	0	0	0	0	-1
4999	1	13.2	6.2	0	0	0	0	-1
5000	1	15.4	5.4	1	1	1	1	1

5001 rows × 8 columns

Algorithms that will be investigated:

This classification project tried 3 different typed of classification algorithms which covered liner regression, like least square regression, LASSO regression and ridge regression; support vectors, like Support Vector Machines; and kernel regression, like kernel ridge regression. For least square regression, LASSO regression and ridge regression also used cross-validation for 5 groups to detect overfitting as well as choose λ for ridge regression and LASSO regression.

Results:

Liner Regression:

- Least Square regression:

  The main idea for least square regression is

  $\hat{y} = Xw$

  $w = \min_{w} || Xw - y||_2^2$

  $w = (X^T X)^{-1} X^T y$

  After using 4001 samples for train, the classifier weight is

```
classifier weight is
[[-0.0288541 ]
 [ 0.05071584]
 [ 0.09615035]
 [ 0.57211291]
 [ 0.51774622]
 [ 0.51989364]
 [ 0.54946535]
 [-2.29150659]]
```

  Using the other 1000 samples for evaluation, error rate is 0.039.

  For using cross-validation in 5 groups, each group has 1000 or 1001 samples, the result for average error rate is 0.051.

  After all, the error rate is less than 6%, which is almost acceptable.

- LASSO regression:

  The main idea for LASSO regression is

  $\hat{y} = Xw$

  $w = \underset{w}{argmin} \, || Xw - y||_2^2 + \lambda ||w||_1$

  Using cross-validation to choose λ from log_{10}^{-6} to log_{10}^{3}.

  After using 4001 samples for train, the classifier weight is

```
classifier weight is
[[-0.0288541 ]
 [ 0.05071584]
 [ 0.09615035]
 [ 0.57211291]
 [ 0.51774622]
 [ 0.51989364]
 [ 0.54946535]
 [-2.29150659]]
```

  Using the other 1000 samples for evaluation, error rate is 0.056.

  For using cross-validation in 5 groups, each group has 1000 or 1001 samples, the result for the average error rate is 0.053.

  After all, the error rate is less than 6%, which is almost acceptable.

- Ridge regression:

  The main idea for Ridge regression is

  $\hat{y} = Xw$

  $w = \underset{w}{argmin} \, || Xw - y||_2^2 + \lambda ||w||_2^2$

  Using cross-validation to choose λ from log_{10}^{-6} to log_{10}^{3}.

  After using 4001 samples for train, the classifier weight is

```
classifier weight is
[[-0.02890307]
 [ 0.05065154]
 [ 0.09603279]
 [ 0.57213185]
 [ 0.5177637 ]
 [ 0.51991086]
 [ 0.54948029]
 [-2.2899463 ]]
```

  Using the other 1000 samples for evaluation, error rate is 0.039.

  For using cross-validation in 5 groups, each group has 1000 or 1001 samples, the result for average error rate is 0.051.

  After all, the error rate is less than 6%, which is almost acceptable.

Support Vector Machines:

The main idea for Support Vector Machines is

$$\hat{y} = Xw$$

$$w = \min_{w}(1 - y_i x_i^T w)_+ + \lambda||\widetilde{w}||_2^2$$

After using 4001 samples for train, the classifier weight is

```
classifier weight is
[[-0.12239776]
 [ 0.17617842]
 [ 0.27060555]
 [ 1.12290397]
 [ 0.99593414]
 [ 1.0082757 ]
 [ 1.03600986]
 [-2.92575872]]
```

Using the other 1000 samples for evaluation, error rate is 0.495.

The error rate is more than 49%, which is not acceptable.

The reason I find is that when the data has more noise, SVM may not perform very well.

Kernel Ridge Regression:

The main idea Kernel Ridge Regression is

$$y(x) = \sum_{i=1}^{N} \alpha_i \, k(x, x_i)$$

Using 4001 samples for train, the weight vectors in kernel space are

```
[-0.05777071 -0.06421154 -0.11128224 ...  0.20535581  0.12056718
 -0.15591126]
```

Using the other 1000 samples for evaluation, error rate is 0.041.

The error rate is less than 5%, which is acceptable.

Strengths and limitations:

The strength for this project is that try to use 3 different type of algorithms, which is 5 different algorithms to classify gender by facial feature. For most of algorithms get acceptable error rate, like least square regression, LASSO regression, ridge regression, and kernel ridge regression. Limitation in this project, SVM shows big error rate. It might be too much noise in

the dataset to make SVM do not perform well. Moreover, the error rate for some of classification is not small enough, which could do some optimization to reduce error in some way like change parameter λ or try other algorithms, like neural network.

Conclusion:

In conclusion, this project using 5 different algorithms in 3 different type, which include linear regression, support vector and kernel regression, to classify gender by facial features. Most of them perform well to get acceptable error rate, which is almost smaller than 5%. However, for SVM shows large error, which may cause by SVM is not good at handle data in large noise. For future work might be focus on find specific reason for the large error of SVM and try more algorithms like neural network.

GitHub:

https://github.com/Mieggy/face-gender-classify