

HW2 Report

Rui Huang, Zonglin Peng

The implementation is written in homework2_rhuang2_zpeng.py

Sample output:

```
> run homework2_rhuang2_zpeng.py
Method  fMSE-training  MSE-testing
Analy.   39.242962989290724      206.79647485479117
GD       83.55009070514743      93.08941828187454
GD reg.  83.56420596173683      93.08632691109096
```

Results and responses:

Full report from code:

Cost fMSEs of the three methods, RMSE of age in the third regressor, and the top 5 most egregious errors images are shown in the report below with their image number in the Dtest, given age y and predicted ages.

MSE cost for all three methods:

Method	MSE-training	MSE-testing
Analytical	39.24296	206.7965
GD w/o reg.	83.55775	93.10225
GD with reg.	83.49984	93.03925

RMSE for GD with regularization:

RMSE of GD with reg.: 13.641059356157214

Note: For RMSE, we use $\sqrt{2 \times fMSE}$ to eliminate the 1/2 in the fMSE, which gives 13.64, otherwise, it will result in: 9.64568557331

Top-5 Errors:

Image Number	y	y_hat
884	10	59.99667
1640	4	52.38498
830	89	41.64049
581	80	33.38617
939	8	53.53574

(The top-5 error images are provided in the same order below in last two pages with visualization of weights)

Conclusion:

As shown in the images below, the method 2 and 3 have clear contours of the face for evaluating the human age by face, where 3 has, and is supposed to have, more characterized weight vectors based on the features on human faces. In this homework, the one-shot analytical weights found the global minimum in the training set but does much worse for the testing set; whereas the gradient descent method has higher cost in training set but does much better in testing set. Therefore, we can say that the analytical overfits the training images.

Extra-credit problem:

As alpha goes to infinity, the GD with regularization will have much worse performance than normal GD method. The reason is that as alpha increases, L2-regularization term cost even higher than the normal gradient term, leading the algorithm want to prioritize minimizing the weights instead of minimizing the cost function. As alpha goes to infinity, the GD with regularization will result in weights approach to zero vector to minimize the cost with the effect of extremely large L2-regularization term.

Here is a sample output for different alpha values:

Alpha = 1000

Method	fMSE-training	fMSE-testing
Analy.	39.242962989290724	206.79647485479117
GD	83.55290113640413	93.10332135998001
GD reg.	89.65256302287754	96.37127424215565

Alpha = 10000

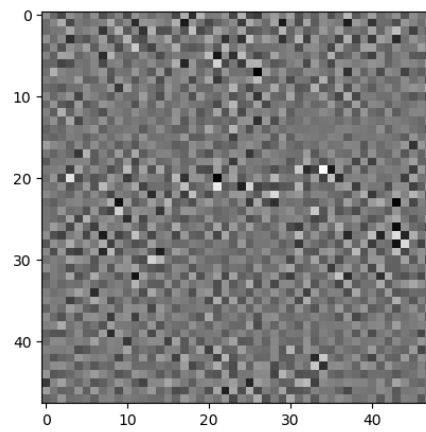
Method	fMSE-training	fMSE-testing
Analy.	39.242962989290724	206.79647485479117
GD	83.5468438162922	93.09532452421172
GD reg.	51043.70114504205	51292.564194081955

Alpha = 100000

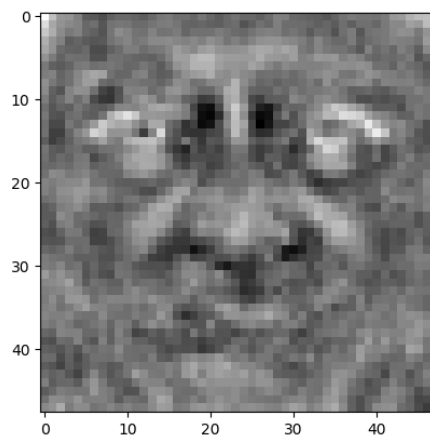
Method	fMSE-training	fMSE-testing
Analy.	39.242962989290724	206.79647485479117
GD	83.55056547609206	93.0863485311576
GD reg.	4.912831714456566e+232	4.95118744625672e+232

Visualization of weights:

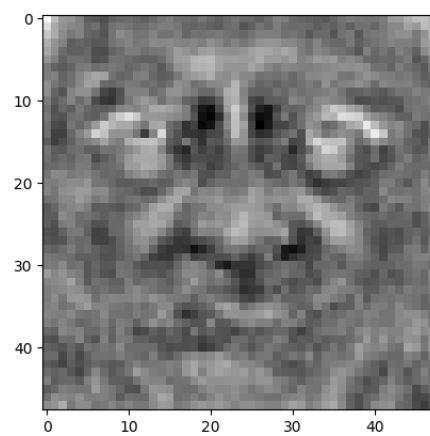
Method 1:



Method 2:

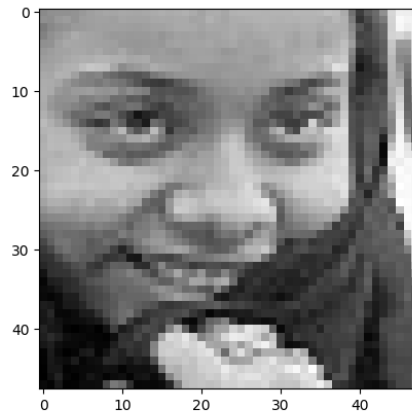


Method 3:

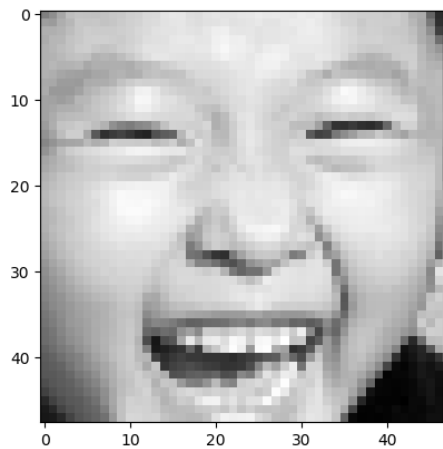


Top-5 most egregious errors images:

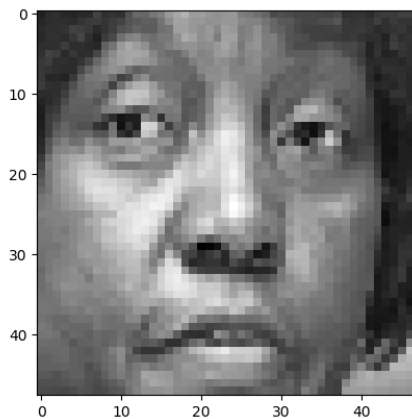
Error Image 1(Number 884):



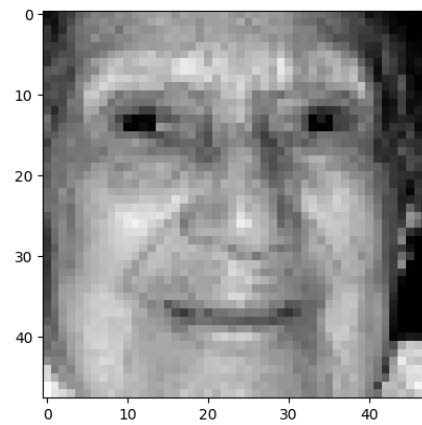
Error Image 2(Number 1640):



Error Image 3 (Number 830):



Error Image 4 (Number 581):



Error Image 5 (Number 939):

