Face Detection and Gender Classification

Leo Pei

Procedure

1. Locate the faces in the picture using OpenCV
2. Principal Components Analysis
3. KNN/SVM/Logistics Regression Classification
4. Accuracy Evaluation/Testing

Data

Data source: 100 face images from Nottingham Face Detection Database

80 Train Data & 20 Test Data

Image Flattening

Each image(350*275) is represented as one row in our table.

Each pixel in the image is tuple containing values for RGB colors.

Each column of the table is flattened R, G, B color for a pixel.

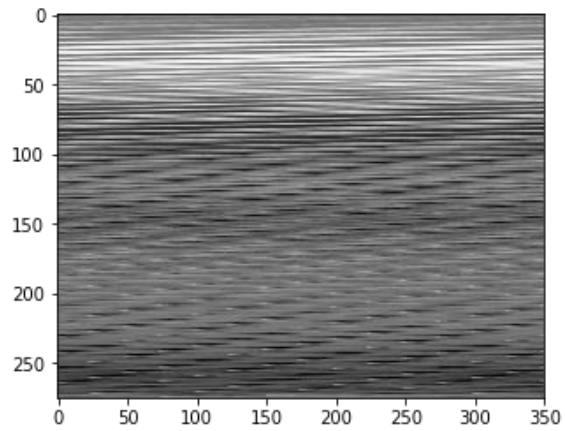
	(0,0) R	(0,0) G	(0,0) B	(0,1) R	(0,1) G	…	(n,n) G	(n,n) B
Pic 1	225	22	74	220	210		0	221
Pic 2	0	12	86	65	225		225	124
Pic 3	123	110	101	99	123		75	50

Principal Component Analysis (PCA)

Each picture has 350*275 = 96,250 pixels.

So each picture has 96,250*3 = **288,750** features.

We used PCA to reduce dimension.

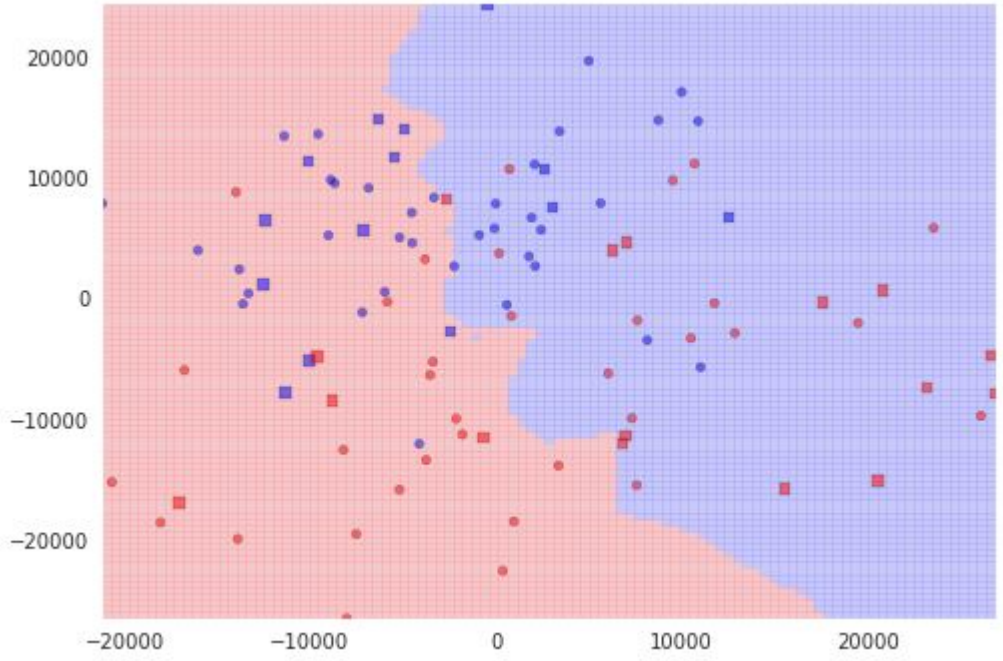


	label	y	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8	pc9	pc10
0	male	True	-4133.844333	11830.209393	-5088.405928	20317.567868	6646.535452	-759.148327	-19389.795051	7246.976677	5085.597174	-9199.953056
1	male	True	2106.072241	-2848.385048	12838.316673	1180.477532	1796.156747	8943.025945	-1488.406058	-2920.470876	-2906.770682	-936.271178
2	male	True	-9026.919095	-5395.428526	3901.475283	916.518500	-10946.540221	-3419.933572	450.757733	-1642.196227	-8862.649040	644.105523
3	male	True	-7204.997278	970.888642	5112.120719	10413.786564	5042.391479	2465.401021	-8066.165393	6020.530894	10673.372563	-418.673500
4	male	True	11008.274266	5499.050785	-11741.958519	16804.263955	7404.427998	-4560.416353	-4343.154309	193.478329	395.432123	-1780.647078

KNN

Accuracy on training data: 0.86

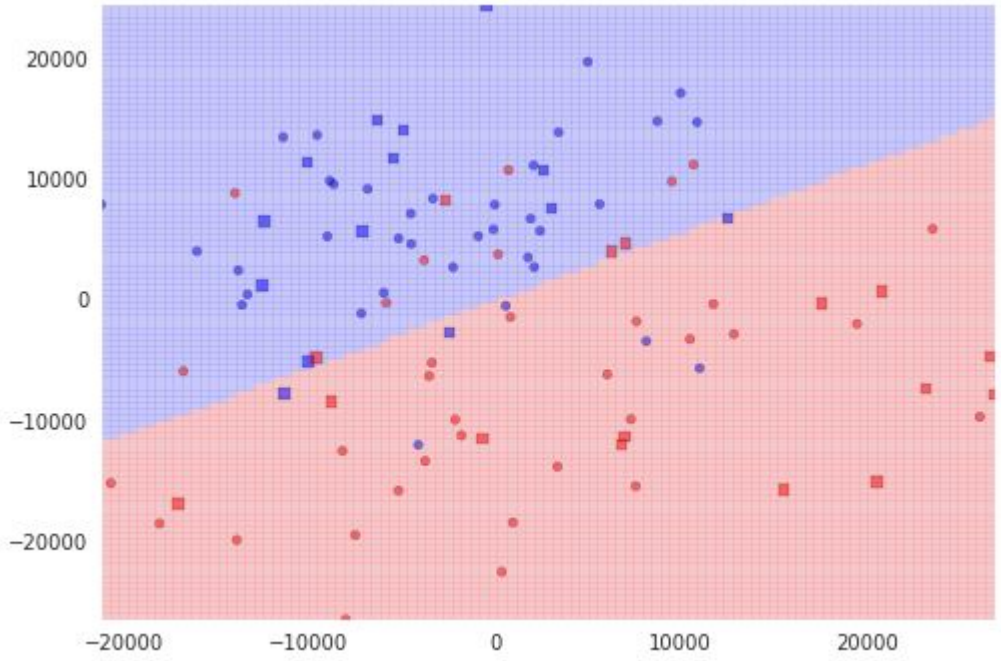
Accuracy on test data: 0.93



Support Vector Machine

Accuracy on training data: 0.89

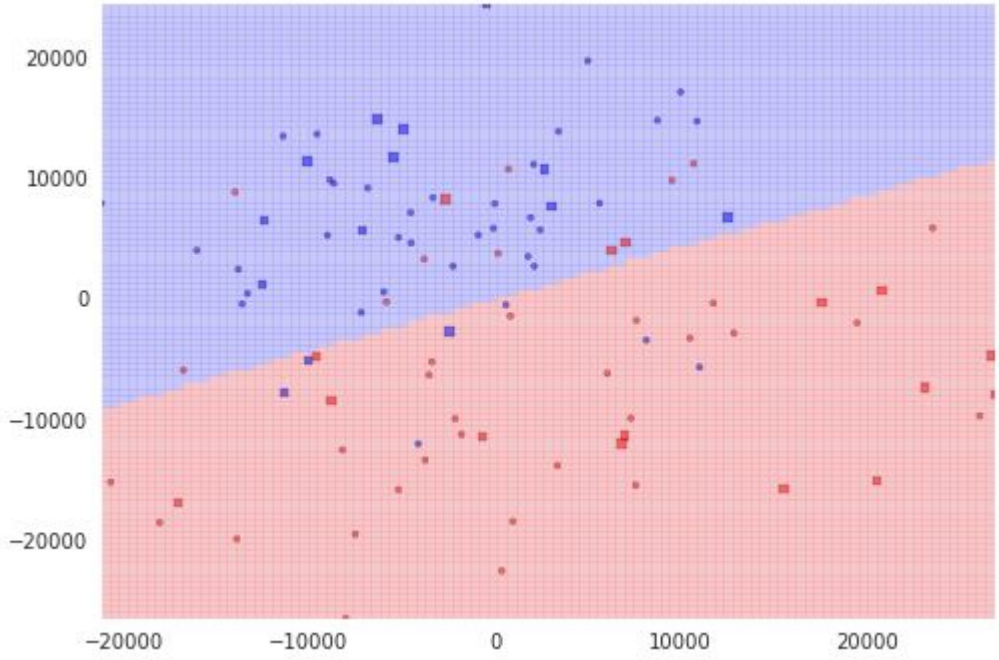
Accuracy on test data: 0.83



Logistic Regression

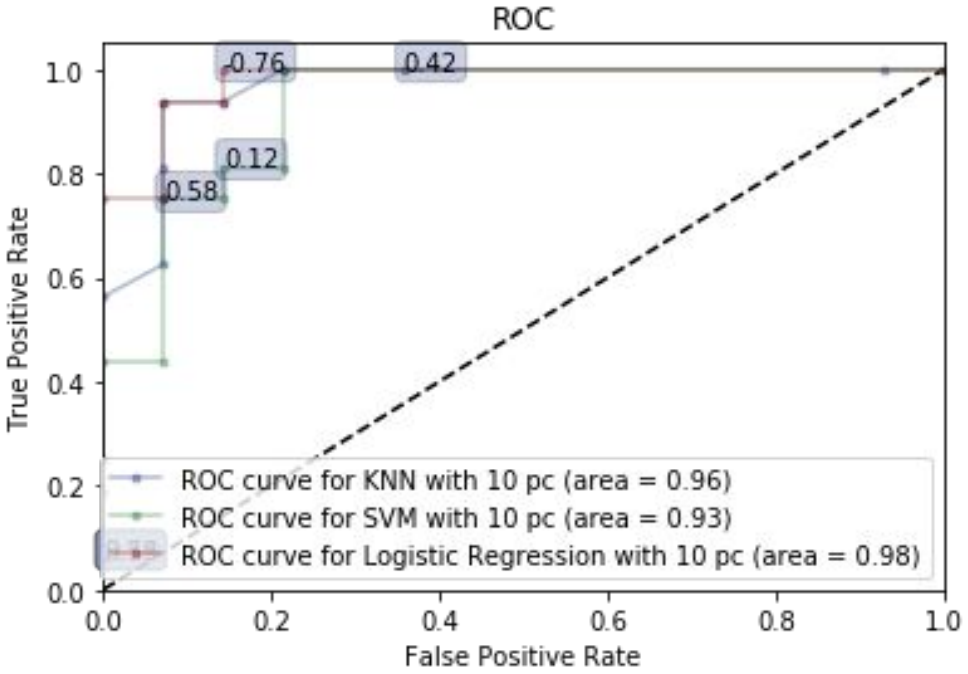
Accuracy on training data: 0.91

Accuracy on test data: 0.90



Model Comparison - ROC Curve

Winner: Logistic Regression



Application