Dimension Reduction

Rachel Huang, Philip Smith

Outline

Project 1:

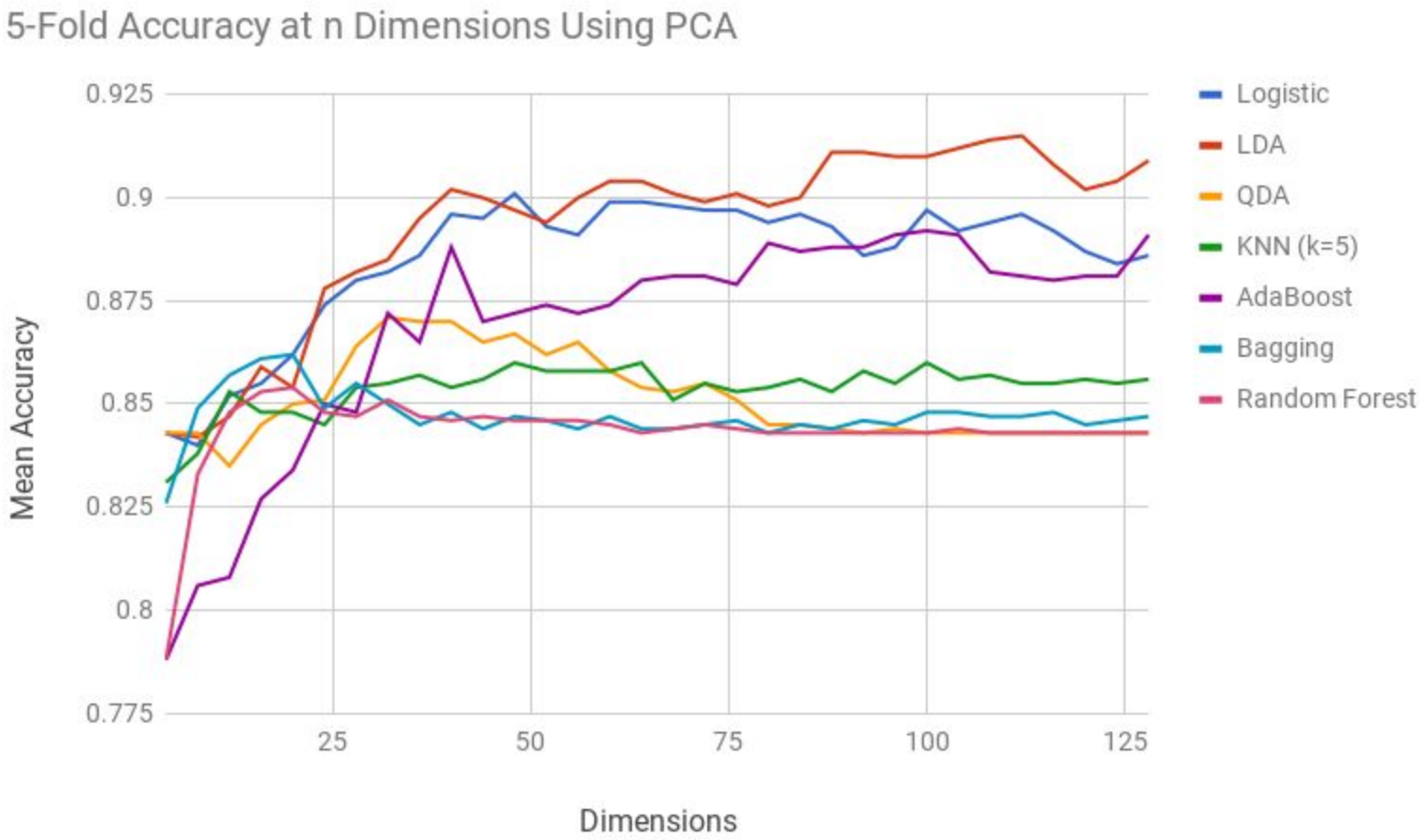
- Conduct PCA on MORPH-II clean data.
- Perform gender classification and age regression.

Project 2:

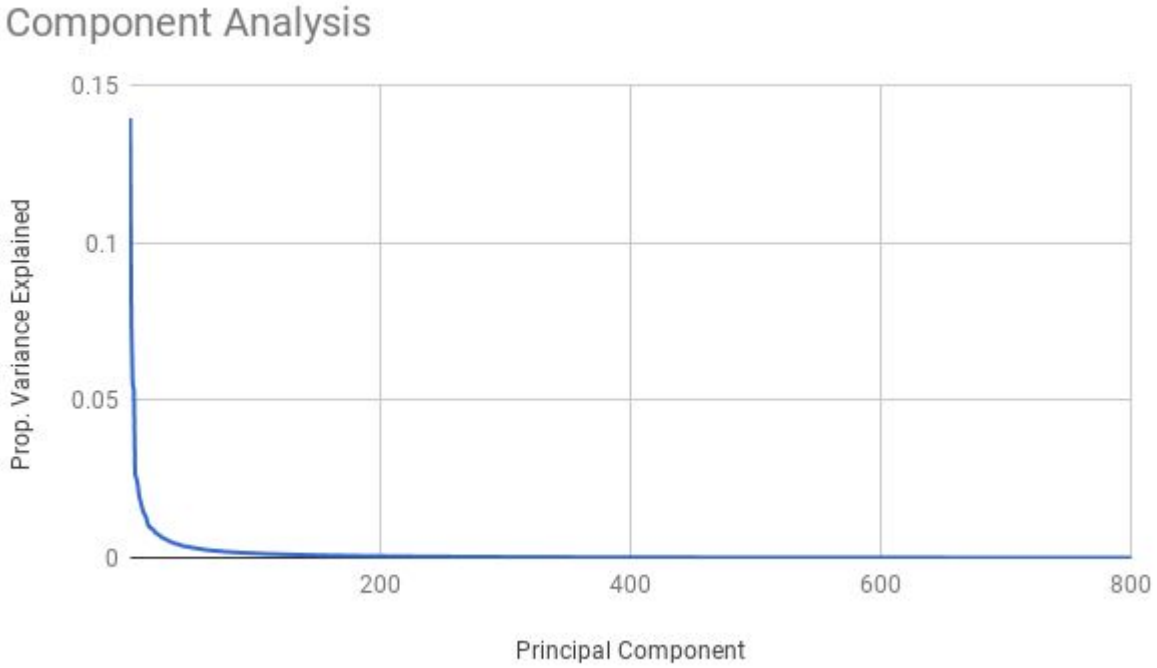
- Conduct Kernel PCA on MORPH-II clean data.
- Perform gender classification and age regression.

Project 3:

- Conduct LDA and KLDA on MORPH-II clean data.
- Perform gender classification and age regression.


5-Fold Accuracy at n Dimensions Using PCA
Mean Accuracy
0.925
0.9
0.875
0.85
0.825
0.8
0.775
25
50
75
100
125
Dimensions
Logistic
LDA
QDA
KNN (k=5)
AdaBoost
Bagging
Random Forest

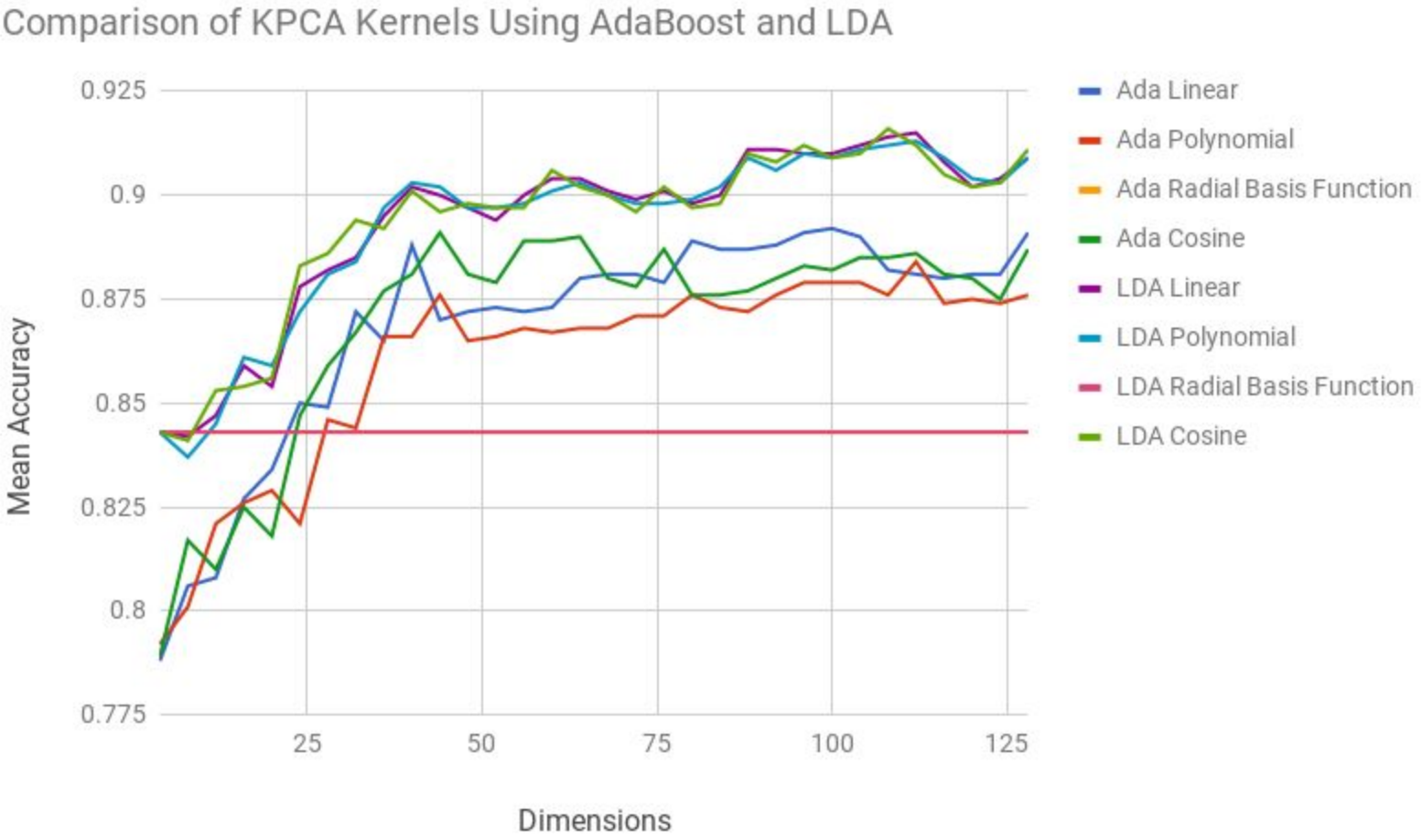
Choose a reasonable number of components



d = 40

Threshold = 0.41%

Explained Variance = 68.3%


Comparison of KPCA Kernels Using AdaBoost and LDA
0.925
0.9
0.875
0.85
0.825
0.8
0.775
Mean Accuracy
25
50
75
100
125
Dimensions
Ada Linear
Ada Polynomial
Ada Radial Basis Function
Ada Cosine
LDA Linear
LDA Polynomial
LDA Radial Basis Function
LDA Cosine

5-Fold CV for Gender

Mean Accuracy	**Logistic Reg.**	**LDA**	**QDA**	**KNN**	**Bagging** trees = 1000	**Random Forest** trees = 1000	**Boosting** trees = 1000	**SVM** (poly)	**SVM** (sigm)
No reduction (*d* = 2568)	**0.909**	0.898	0.569	0.860	0.866	0.862	**0.904**	0.871	0.782
PCA (*d* = 40)	0.896	**0.902**	0.87	0.854	0.848	0.846	0.888	**0.891**	0.843
KPCA: Cosine Kernel (*d* = 40)	0.843	0.901	0.871	0.864	0.853	0.846	0.881	0.843	0.843
LDA (*d* = 1)	0.88	———	**0.877**	**0.882**	**0.883**	**0.883**	0.883	0.885	0.624

Gender Misclassification (LDA w/ PCA)

ID: 061370_3F55

Predicted: Male

Probability: 99%

Actual: Female

ID: 065790_2M47

Predicted: Female

Probability: 80%

Actual: Male

5-Fold CV for Age

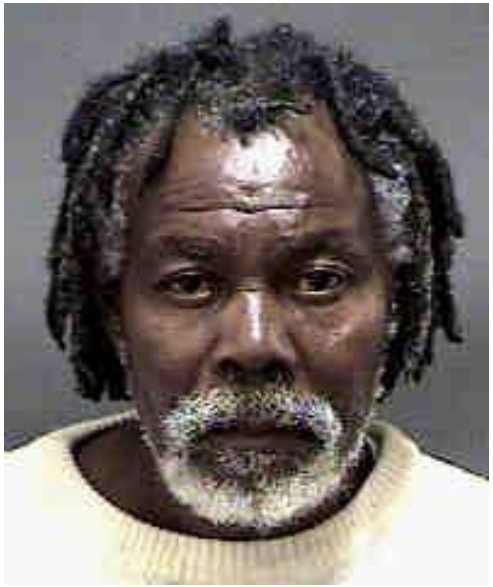
MAE	**Bagging** trees = 1000	**Random Forest** trees = 1000	**Boosting** trees = 1000	**SVM** (polynomial)	**SVM** (sigmoid)
No reduction (*d* = 2568)	10.86	10.11	13.54	10.001	17.804
PCA (*d* = 40)	10.97	10.99	15.42	11.94	12.28
KPCA (*d* = 40)	9.932	**9.719**	11.781	14.258	14.189
LDA (*d* = 40)	**9.89**	10.416	**9.07**	**9.071**	**9.07**

Random Forest Age Predictions (PCA)

ID: 065844_1F46

Predicted: 40

Actual: 46



ID: 046297_5M56

Predicted: 17

Actual: 56

Questions?