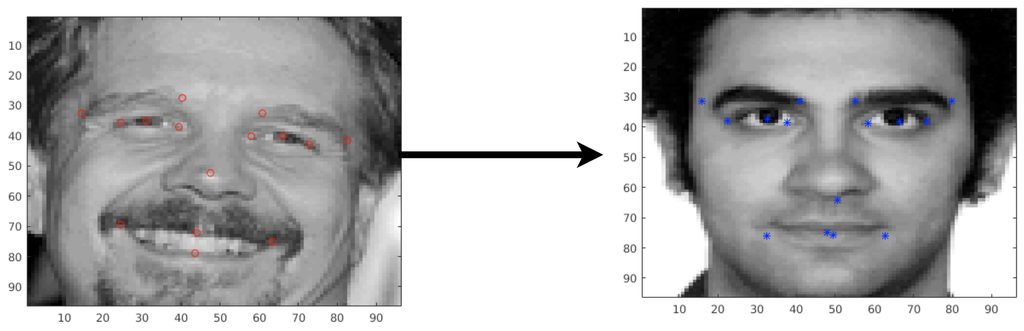
Facial Keypoint Detection

Ashkan Esmaeili, Khashayar Khosravi, Shahab Mirjalili
CS 229 Project
Stanford University
Fall 2015

Introduction and Application

Predicting location of 15 keypoints on a given image after learning these points on the training set



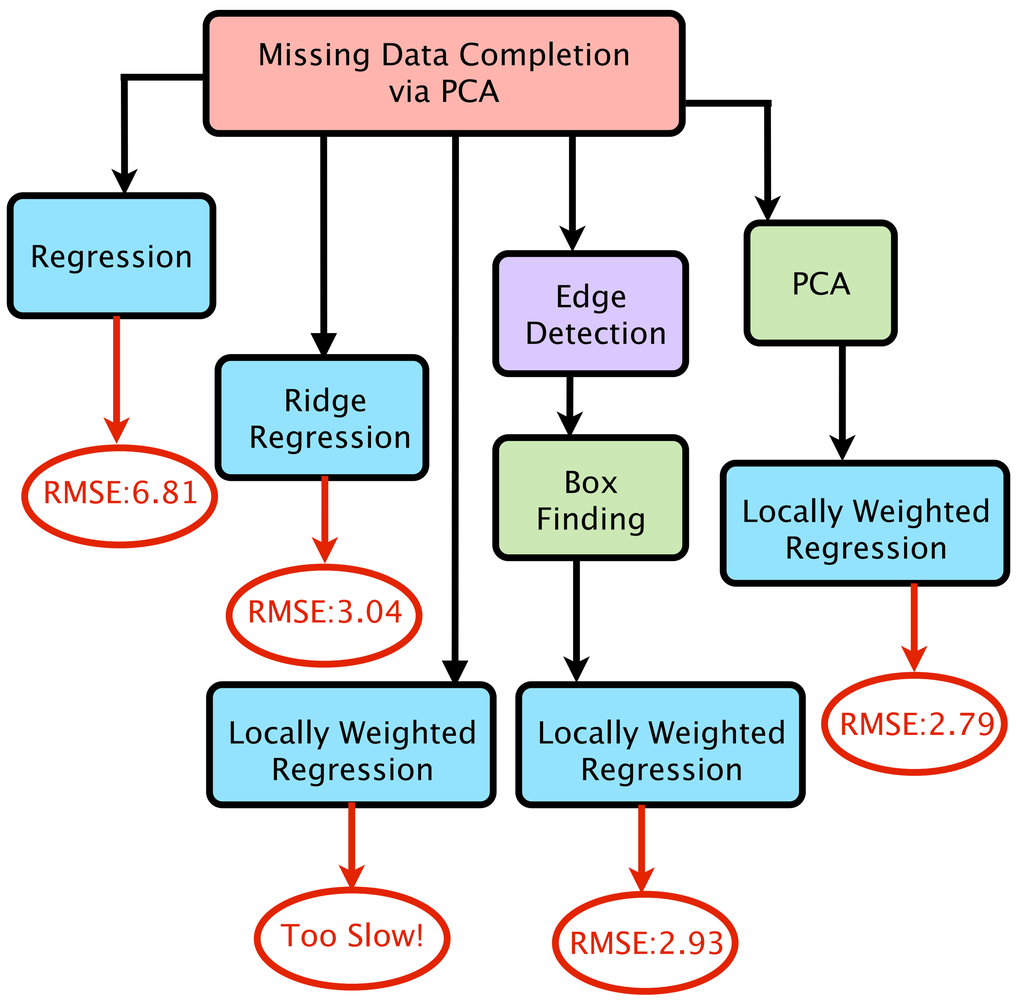
7049 training images

arbitrary face photo as test

Applications:

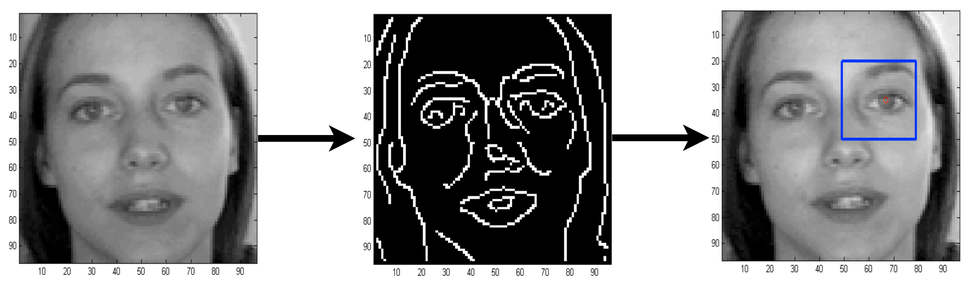
- **Tracking faces in images and videos**
- **Analysing facial expressions**
- **Biometrics / face recognition**

Algorithms



Notes

- **Only ~2000/~7000 of the photos have no missing target variables, making matrix completion a necessary step of prediction**
- **Regression naive and prone to overfitting since number of features is larger than training set size ==> LWLR and Feature selection**
- **LWLR too expensive and impractical**
- **Using the edge detected versions of the photos, we find boxes around keypoints and use the smaller box photos in predictions**



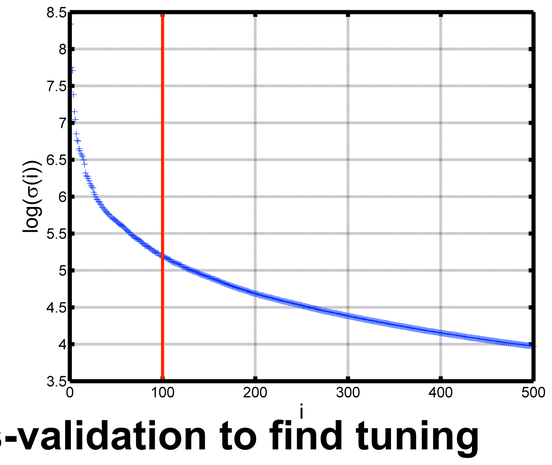
- **After dimension reduction via PCA, we use locally weighted linear regression to predict keypoints**

PCLWR

PCA dimension of k=100:
- Maintains accuracy
- Computationally manageable for weighted linear regression:

$$X = U\Sigma V^T$$

$$\tilde{x}^{(i)} = \begin{bmatrix} v_1^T x^{(i)} \\ v_2^T x^{(i)} \\ \vdots \\ v_k^T x^{(i)} \end{bmatrix} \in R^k$$



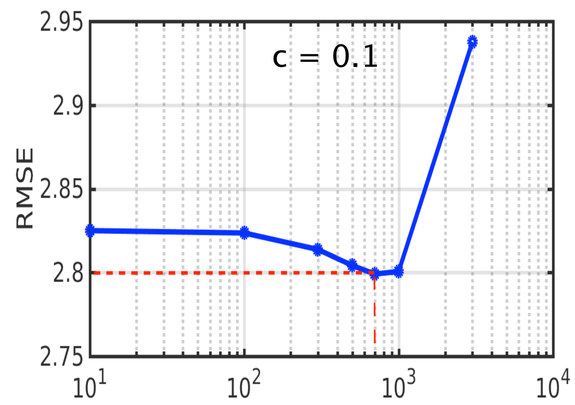
Using hold-out cross-validation to find tuning parameters for weighted linear regression:

$$D_i = \| x^{(\tilde{test})} - x^{(\tilde{i})} \|^2$$

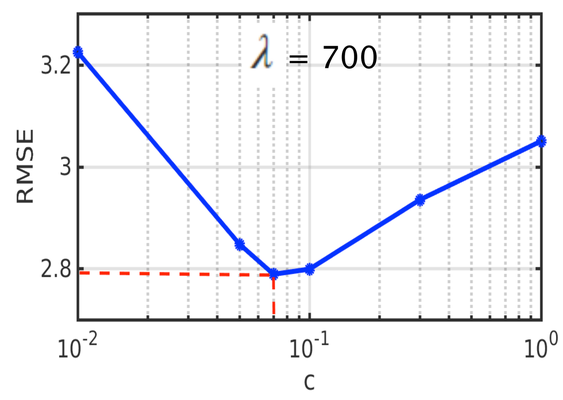
$$W_i = exp(-\frac{D_i}{(2c)max(D)})$$

$$\theta = (\tilde{X}^T W \tilde{X} + \lambda I)^{-1} \tilde{X}^T W y$$

$$y^{(\hat{test})} = \theta^T x^{(\tilde{test})}$$



$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \tilde{y}^{(i)})^2}$$

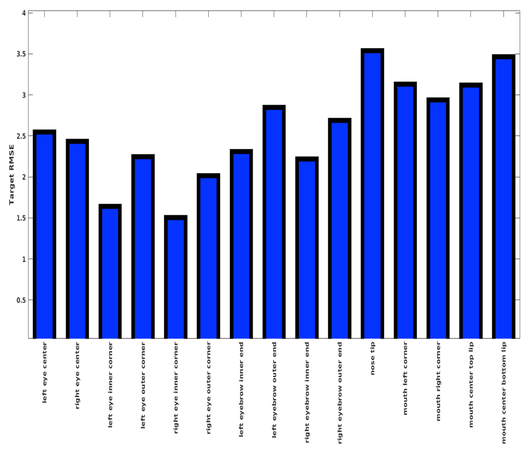


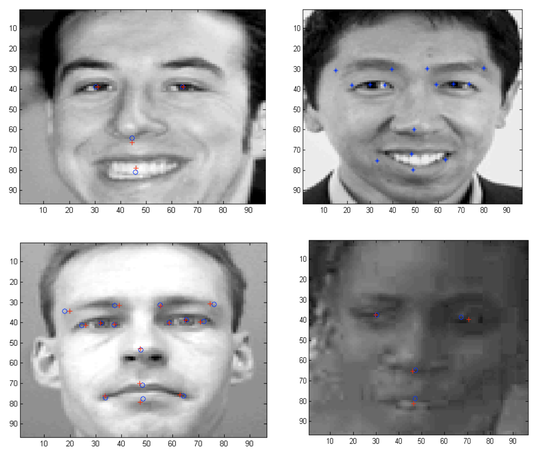
Chosen parameters:

λ = 700 c = 0.07

RMSE:2.79

Results





Summary

- **Explored different methods for dimension reduction and predicting keypoint values on 96x96 face photos**
- **Optimized tuning parameters for each method based on hold-out cross-validation**
- **Employing locally weighted regression on the dimensionally reduced features obtained from PCA yielded least errors on the target values**
- **Root mean squared error for all target variables was 2.79 pixels**