2413, Machine Learning, Homework 5
Due Date: 11/12/2013
Universität Bern

Question 1 (1 points) Provide a function that performs the normalization of the data matrix in the attached face classification problem. Such function should:

- Compute $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$.
- Replace each $x^{(i)}$ with $x^{(i)} - \mu$.
- Let $\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)})^2$.
- Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$.

The function specification is the following:

```
function [NormData] = normalizeData_YourName(Data)
% Input
%     Data: the data matrix, each column represents a data point.
%
% Output
%     NormData: the normalized data matrix.
%
% Your code here
```

Substitute "YourName" with your name and surname in capital letters. Add comments in your code to explain the most important parts of your algorithm.

Question 2 (1 points) Provide a function that implements the PCA data dimensionality reduction algorithm. You are only allowed to use the build in function *svd*. The function specification is the following:

```
function Eigen = pca_YourName(Data, k)
% Input
%     Data: the data matrix, each column represents a data point.
%     k: number of principle components to use.
%
% Output
%     Eigen: a matrix with k columns, where each column is an
%                eigenvector of the covariance matrix of Data.
%
% Your code here
```

Substitute "YourName" with your name and surname in capital letters. Add comments in your code to explain the most important parts of your algorithm.

Question 3 (1 points) Provide a function that projects the training and testing data on the PCA basis computed in the previous step.

The function specification is the following:

```
function Projected = project_YourName(Data, Basis)
% Input
%     Data: the data matrix, each column represents a data point.
%     Basis: the PCA basis.
%
% Output
%     Projected: the projected data.
%
% Your code here
```

Substitute "YourName" with your name and surname in capital letters. Add comments in your code to explain the most important parts of your algorithm.

Question 4 (2 points) Plot the average accuracy of the classifier as a function of the number of principal components k that form the PCA basis. To compute the average accuracy run the SVM classifier 10 times for each value of k. To avoid long computations you can perform this test every 5 k, therefore you will evaluate the classifier for the values $k = \{1, 6, 11, \ldots, 576\}$.

Find the minimum number of principal components that give an accuracy of at least 90 % on average.