

# Homework 4

## Problem. 1. (20 points)

Download the “teapots.mat” data set containing 100 images of teapots of size  $38 \times 50$ . To view an image, say the second one in the data set type:

```
imagesc(reshape(teapotImages(2,:),38,50));  
colormap gray;
```

Compute the data mean and top 3 eigenvectors of the data covariance matrix and show them as images. Reconstruct the data using PCA with least squares error using only the mean and a linear combination of the top 3 eigenvectors. Show 10 different images before and after reconstruction. Discuss results.

## Problem. 2. (10 points)

Suppose we have a box containing \$8\$ apples and \$4\$ oranges and a second box containing \$10\$ apples and \$2\$ oranges. One of the boxes is chosen at random (with equal probability of choosing either box) and an item is selected from the box and found to be an apple. Use Bayes' rule to find the probability that the apple came from the first box.

## Problem. 3. (20 points)

Suppose you have a 2-class classification problem, where each class is Gaussian. Let  $\theta = \{\alpha, \mu_1, \Sigma_1, \mu_2, \Sigma_2\}$  denote the set of model parameters. Suppose the class probability  $p(y|\theta)$  is modelled via the Bernoulli distribution, i.e.  $p(y|\theta) = \alpha^y(1 - \alpha)^{1-y}$ , and the probability of the data  $p(x|y, \theta)$  is modelled as  $p(x|y, \theta) = \mathcal{N}(x|\mu_y, \Sigma_y)$ . Recover the parameters of the model from maximum likelihood approach.

Assume the data are i.i.d. and  $N$  is the number of data samples. Show all derivations. Next, suppose you want to make a classification decision by assigning proper label  $y$  to a given data point  $x$ . You decide the label based on Bayes optimal decision  $y = \arg \max_{\hat{y} \in \{0,1\}} p(\hat{y}|x)$ . Prove that the decision boundary is linear when covariances  $\Sigma_1$  and  $\Sigma_2$  are equal and otherwise the boundary is quadratic.

Illustrate on 2d example the decision boundary for the case when covariances are not equal clearly indicating which class is more concentrated around its mean.