CSE 555 Pattern Recognition HW2 Reference Solution

Albert Y. C. Chen

Computer Science and Engineering
University at Buffalo, The State University of New York

Assigned: 02/05/2009
Due: 02/24/2009

Problem 1 (20%)

1. The prior model $P(\omega)$ constrained by c parameters $\theta = (\theta_1, \theta_2, ..., \theta_c)$ can be thought as c priors each constrained by one of the θ_i parameter: $P(\omega_i) = \theta_i$, where $i = 1, ..., c$. Note that "Prior probabilities" are not based on the features x_i of the samples but the frequency of occurrence for each class, a new variable s_{ik} is introduced for the MLE estimation: $s_{ik} = 1$ if the k^{th} sample belongs to class ω_i and $s_{ik} = 0$ otherwise. For any sample x_j, the probability of it belonging to class ω_i is θ_i and $1 - \theta_i$ if not, which can be written as:

$$P(s_{ik}|\theta_i) = (\theta_i)^{s_{ik}}(1-\theta_i)^{1-s_{ik}} \quad .$$

The joint probability density function for the entire training data set is:

$$P(\mathcal{D}|\theta_i) = P(s_{i1}, ..., s_{iN}|\theta_i) = \prod_{k=1}^{N} (\theta_i)^{s_{ik}}(1-\theta_i)^{1-s_{ik}} \quad .$$

The log-likelihood function is therefore:

$$\begin{aligned} l(\theta_i) \equiv \ln P(\mathcal{D}|\theta_i) &= \ln \Big[\prod_{k=1}^{N} (\theta_i)^{s_{ik}}(1-\theta_i)^{1-s_{ik}} \Big] = \sum_{k=1}^{N} \ln(\theta_i)^{s_{ik}} + \sum_{k=1}^{N} \ln(1-\theta_i)^{1-s_{ik}} \\ &= \sum_{k=1}^{N} s_{ik} \ln \theta_i + \sum_{k=1}^{N} (1-s_{ik}) \ln(1-\theta_i) \\ &= (\ln \theta_i) \sum_{k=1}^{N} s_{ik} + (\ln(1-\theta_i)) \sum_{k=1}^{N} (1-s_{ik}) \\ &= n_i \ln \theta_i + (N-n_i) \ln(1-\theta_i) \quad . \end{aligned}$$

2. The above log-likelihood function is maximized at points where its derivative is 0:

$$\frac{n_i}{\theta_i} + \frac{N-n_i}{\theta_i - 1} = 0 \quad \Rightarrow \quad \frac{n_i(\theta_i - 1) + (N-n_i)\theta_i}{\theta_i(\theta_i - 1)} = 0 \quad \Rightarrow \quad \frac{-n_i + N\theta_i}{\theta_i(\theta_i - 1)} = 0$$

$$\Rightarrow \quad N\theta_i = n_i \quad \Rightarrow \quad \theta_i = \frac{n_i}{N} \quad \text{, thus proved!}$$

Problem 2 (15%)

As discussed in class on 2/24/2009.

Problem 3 (20%)

1. For $d = 1$, the "volume" r of this one dimension hypercube is simply the length of a line, therefore the length of the "side" of this "volume" $l = r$.
   For $d = 2$, the "volume" of this two dimensional hypercube is the area of a square, therefore $l^2 = r$, $l = r^{1/2}$.

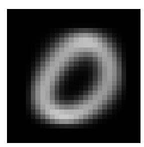
2. For a d dimensional hypercube $l^d = r$, thus $l = r^{1/d}$. When $d = 100$ and $r = 0.1$, we have $l = 0.1^{1/100} \approx 0.9772$, textiti.e. the length of each side of the hypercubic neighborhood is 0.977, almost as long as the length of each side of the unit hypercube.

3. $0.01^{1/10100} = 0.9995$
   $0.1^{1/10100} = 0.9998$.

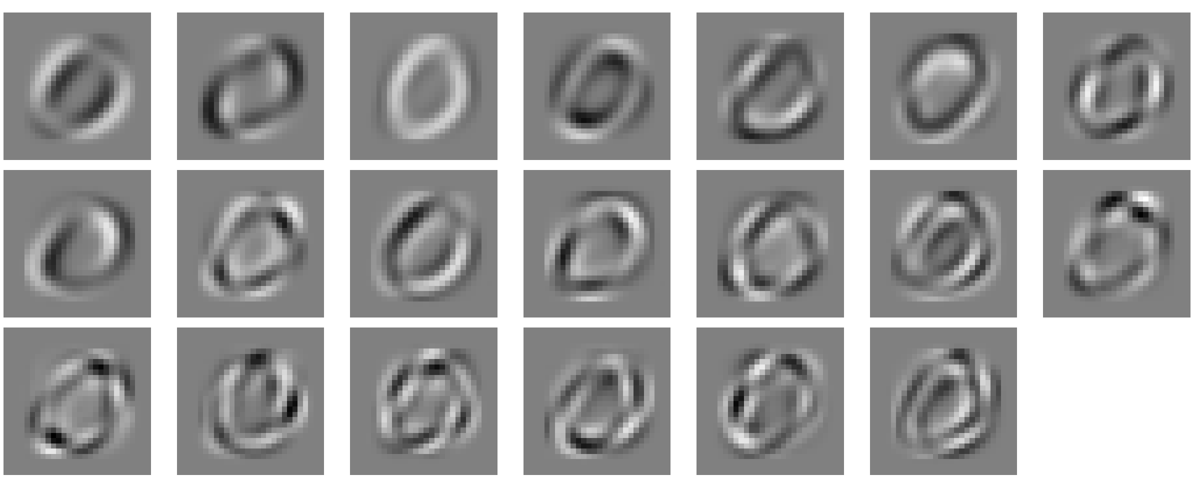
Problem 4 (45%)

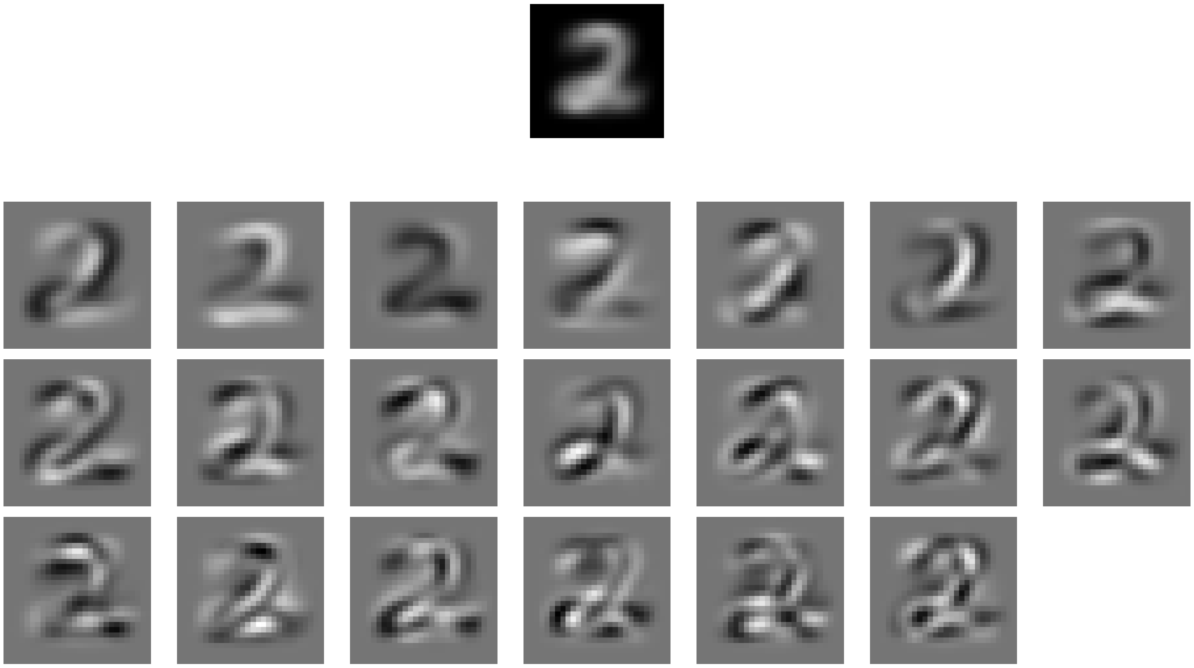
1. Please refer to the sample code.

2. The mean image of digit 0:



   The first 20 eigenvectors as images for the digit 0 (from left to right, from top to bottom):





3. The mean image of digit 2 and the first 20 eigenvectors as images for the digit 2 (from left to right, from top to bottom):
4. Omitted
5. Omitted
6. Omitted