UNIVERSITY OF SOUTHERN CALIFORNIA

Homework 5

Peehoo Dewan
Collaberated with Coco Chengliangdong, Pooja Voladoddi, Satakshi Rana

December 3, 2014

1 PRINCIPAL COMPONENT ANALYSIS

1.1 DERIVATION IN TERMS OF RECONSTRUCTION ERROR

(a) Given,

$$\text{reconstruction error} = \mathrm{E} = \sum_{i=1}^{N} ||\mathbf{x}_i - \hat{\mathbf{x}}_i||_2^2 \tag{1.1}$$

We can reconstruct $\mathbf{x}_i$ by substituting $\hat{\mathbf{x}}_i = \mathbf{U}\mathbf{z}_i$
Substituting the above in equation 1.1 we get,

$$\begin{aligned} E &= \sum_{i=1}^{N} ||\mathbf{x}_i - \mathbf{U}\mathbf{z}_i||_2^2 \\ &= \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{U}\mathbf{z}_i)^T(\mathbf{x}_i - \mathbf{U}\mathbf{z}_i) \\ &= \sum_{i=1}^{N} (\mathbf{x}_i^T\mathbf{x}_i - 2(\mathbf{U}\mathbf{z}_i)^T\mathbf{x}_i + (\mathbf{U}\mathbf{z}_i)^T\mathbf{U}\mathbf{z}_i) \end{aligned}$$

Differentiating w.r.t $\mathbf{z}_i$ and setting to 0 gives,

$$\frac{\partial E}{\partial \mathbf{z}_i} = (-2)\mathbf{U}^T\mathbf{x}_i + 2\mathbf{U}^T\mathbf{U}\mathbf{z}_i := 0$$

Since $\mathbf{U}^T\mathbf{U} = \mathbf{I}_d$, the optimal value of $\mathbf{z}_i$ which minimizes reconstruction error is,

$$\mathbf{z}_i = \mathbf{U}^T\mathbf{x}_i$$

(b) Optimal U for PCA solution Given,

$$E = \sum_{i=1}^{N} ||\mathbf{x}_i - \hat{\mathbf{x}}_i||_2^2$$

We can reconstruct $\mathbf{x}_i$ by substituting $\hat{\mathbf{x}_i} = \mathbf{U}\mathbf{z}_i$ and since we derived above that the optimal value of $\mathbf{z}_i = \mathbf{U}^T\mathbf{x}_i$, therefore we get $\hat{\mathbf{x}_i} = \mathbf{U}\mathbf{U}^T\mathbf{x}_i$. We are also given the constraint that $\mathbf{U}^T\mathbf{U} = \mathbf{I}_d$. So we will solve this problem with lagragian as below,

$$\begin{aligned} L(\lambda, \mathbf{U}) &= \sum_{i=1}^{N} ||\mathbf{x}_i - \mathbf{U}\mathbf{U}^T\mathbf{x}_i||_2^2 + \lambda(\mathbf{U}^T\mathbf{U} - \mathbf{I}_d) \\ &= \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{U}\mathbf{U}^T\mathbf{x}_i)^T(\mathbf{x}_i - \mathbf{U}\mathbf{U}^T\mathbf{x}_i) + \lambda(\mathbf{U}^T\mathbf{U} - \mathbf{I}_d) \end{aligned}$$

Differentiating w.r.t $\mathbf{U}$ we get,

$$\frac{\partial L(\lambda, \mathbf{U})}{\partial \mathbf{U}} = \sum_{i=1}^{N} (-2)\mathbf{x}_i\mathbf{x}_i^T\mathbf{U} + 2\lambda\mathbf{U} := 0$$

Solving the above gives,

$$\mathbf{X}^T\mathbf{X}\mathbf{U} = \lambda\mathbf{U}$$

The above result shows that $\mathbf{U}$ contains the set of eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$

1.2 Projecting a Gaussian distribution

(a)Given,

$$\mathbf{x} \sim N(0, \Sigma)$$

Since we project x into one dimensional space such that $\mathbf{z} = \mathbf{p}^T\mathbf{x}$. Therefore we can write that,

$$\mathbf{z} \sim N(0, \mathbf{p}^T\Sigma\mathbf{p})$$

We can write the entropy of the gaussian as,

$$H(\mathbf{z}) = -\int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} = \frac{1}{2}(1 + \ln 2\pi) + \frac{1}{2}\ln(\mathbf{p}^T\Sigma\mathbf{p})$$

In order to find the optimal $\mathbf{p}$ that maximizes entropy, we have to maximize $H(\mathbf{z})$. The first half of the $H(\mathbf{z})$ is constant and therfore independent of $\mathbf{p}$. So in order to maximize $H(\mathbf{z})$, we maximize the second term. We also see that logarithm is an increasing function, therefore maximizing $\ln(\mathbf{p}^T\Sigma\mathbf{p})$ is same as maximizing $\mathbf{p}^T\Sigma\mathbf{p}$. We are also given a constraint that $\mathbf{p}^T\mathbf{p} = 1$. Therefore solving using lagragian multiplier we get,

$$L(\lambda, \mathbf{p}) = \mathbf{p}^T\Sigma\mathbf{p} - \lambda(\mathbf{p^T}\mathbf{p} - \mathbf{1})$$

Differentiating w.r.t **p** we get,

$$\frac{\partial L(\lambda, \mathbf{p})}{\partial \mathbf{p}} = 2\Sigma\mathbf{p} - \mathbf{2}\lambda\mathbf{p} := \mathbf{0}$$

Solving the above gives,

$$\Sigma\mathbf{p}^* = \lambda\mathbf{p}^*$$

(b) We see that the above solution gives **p** as the eigenvectors of the covariance matrix Σ. We have seen in lecture and also in the previous question that this solution is also the solution of PCA and therefore minimizes reconstruction error or maximizes variance.

2 Hidden Markov model

2.1 Probability of the sequence

Given,

$$\text{Observed Sequence} = e = CGTCAG$$

We define

$$\alpha_t(i) = P(Q_t = S_i, O_1, \ldots, O_t)$$

If the length of the sequence is T and the HMM has transition matrix of A and output matrix of B, then the initialization and iteration stepps are as below:

Initialization

$$\alpha_1(i) = \pi(i) * B_i(o_1) \text{ for every i (that is every node} S_i)$$

Iteration step

$$\alpha_{t+1}(i) = (\sum_{j=1}^{N} \alpha_t(j) a_{j,i}) * b_i(o_{t+1}) \text{ for every i}$$

Plugging in the above values, we can calculate all the values of α's via forward propagation as below:

$$\begin{aligned}
\alpha_1(1) &= P(Q_1 = S_1)P(O_1 = C|Q_1 = S_1) \\
&= \pi_1 b_1(C) \\
&= 0.5 \times 0.1 = 0.05 \\
\alpha_1(2) &= P(Q_1 = S_2)P(O_1 = C|Q_1 = S_2) \\
&= \pi_2 b_2(C) \\
&= 0.5 \times 0.4 = 0.20
\end{aligned}$$

Similarly,

$$\alpha_2(1) = (\alpha_1(1)a_{11} + \alpha_1(2)a_{21})b_1(G) = (0.05 \times 0.7 + 0.2 \times 0.3) \times 0.4 = 0.038$$
$$\alpha_2(2) = (\alpha_1(1)a_{12} + \alpha_1(2)a_{22})b_2(G) = (0.05 \times 0.3 + 0.2 \times 0.7) \times 0.1 = 0.0155$$
$$\alpha_3(1) = (\alpha_2(1)a_{11} + \alpha_2(2)a_{21})b_1(T) = (0.038 \times 0.7 + 0.0155 \times 0.3) \times 0.1 = 0.003125$$
$$\alpha_3(2) = (\alpha_2(1)a_{12} + \alpha_2(2)a_{22})b_2(T) = (0.038 \times 0.3 + 0.0155 \times 0.7) \times 0.4 = 0.0089$$
$$\alpha_4(1) = (\alpha_3(1)a_{11} + \alpha_3(2)a_{21})b_1(C) = (0.003125 \times 0.7 + 0.0089 \times 0.3) \times 0.1 = 0.00048575$$
$$\alpha_4(2) = (\alpha_3(1)a_{12} + \alpha_3(2)a_{22})b_2(C) = (0.003125 \times 0.3 + 0.0089 \times 0.7) \times 0.4 = 0.002867$$
$$\alpha_5(1) = (\alpha_4(1)a_{11} + \alpha_4(2)a_{21})b_1(A) = (0.00048575 \times 0.7 + 0.002867 \times 0.3) \times 0.4 = 0.00048005$$
$$\alpha_5(2) = (\alpha_4(1)a_{12} + \alpha_4(2)a_{22})b_2(A) = (0.00048575 \times 0.3 + 0.002867 \times 0.7) \times 0.1 = 0.0002152625$$
$$\alpha_6(1) = (\alpha_5(1)a_{11} + \alpha_5(2)a_{21})b_1(G) = (0.00048005 \times 0.7 + 0.0002152625 \times 0.3) \times 0.4 = 0.0001602455$$
$$\alpha_6(2) = (\alpha_5(1)a_{12} + \alpha_5(2)a_{22})b_2(G) = (0.00048005 \times 0.3 + 0.0002152625 \times 0.7) \times 0.1 = 0.00002946987$$

Therefore probability of the sequence e is calculated as:

$$\begin{aligned} P(e|\theta) &= \alpha_6(1) + \alpha_6(2) \\ &= 0.0001897154 \end{aligned}$$

2.2 PROBABILITY FOR HIDDEN STATES

In order to calculate this, we firts have to calculate the backward probabilities: We define

$$\beta_t(i) = P(O_{t+1}, \ldots, O_T | Q_t = S_i,)$$

Initialization

$$\beta_T(i) = P(\emptyset | S_i) = 1$$

Iteration step

$$\beta_t(i) = \sum_{j=1}^{N} a_{j,i} P(o_{t+1}|S_j) * \beta_{t+1}(j)$$

Plugging in the above values, we can calculate all the values of β's via backward propagation as below:

$\beta_6(1) = 1$

$\beta_6(2) = 1$

$\beta_5(1) = \beta_6(1)a_{11}b_1(G) + \beta_6(2)a_{12}b_2(G) = (0.7 \times 0.4 \times 1 + 0.3 \times 0.1 \times 1) = 0.31$

$\beta_5(2) = \beta_6(1)a_{21}b_1(G) + \beta_6(2)a_{22}b_2(G) = (0.3 \times 0.4 \times 1 + 0.7 \times 0.1 \times 1) = 0.19$

$\beta_4(1) = \beta_5(1)a_{11}b_1(A) + \beta_5(2)a_{12}b_2(A) = (0.7 \times 0.4 \times 0.31 + 0.3 \times 0.1 \times 0.19) = 0.0925$

$\beta_4(2) = \beta_5(1)a_{21}b_1(A) + \beta_5(2)a_{22}b_2(A) = (0.3 \times 0.4 \times 0.31 + 0.7 \times 0.1 \times 0.19) = 0.0505$

$\beta_3(1) = \beta_4(1)a_{11}b_1(C) + \beta_4(2)a_{12}b_2(C) = (0.7 \times 0.1 \times 0.0925 + 0.3 \times 0.4 \times 0.0505) = 0.012535$

$\beta_3(2) = \beta_4(1)a_{21}b_1(C) + \beta_4(2)a_{22}b_2(C) = (0.3 \times 0.1 \times 0.0925 + 0.7 \times 0.4 \times 0.0505) = 0.016915$

$\beta_2(1) = \beta_3(1)a_{11}b_1(T) + \beta_3(2)a_{12}b_2(T) = (0.7 \times 0.1 \times 0.012535 + 0.3 \times 0.4 \times 0.016915) = 0.00290725$

$\beta_2(2) = \beta_3(1)a_{21}b_1(T) + \beta_3(2)a_{22}b_2(T) = (0.3 \times 0.1 \times 0.012535 + 0.7 \times 0.4 \times 0.016915) = 0.00511225$

$\beta_1(1) = \beta_2(1)a_{11}b_1(G) + \beta_2(2)a_{12}b_2(G) = (0.7 \times 0.4 \times 0.00290725 + 0.3 \times 0.1 \times 0.00511225) = 0.0009673975$

$\beta_1(2) = \beta_2(1)a_{21}b_1(G) + \beta_2(2)a_{22}b_2(G) = (0.3 \times 0.4 \times 0.00290725 + 0.7 \times 0.1 \times 0.00511225) = 0.0007067275$

We can now define that we calculate the probability of the hidden states by the following formula:

$$P(q_t = S_i | o_1 \ldots o_T) = \frac{\alpha_t(i) * \beta_t(i)}{\alpha_6(1) + \alpha_6(2)} \text{ at time t}$$

Substituting in the above formula, we get the following values:

$$P(X_1 = S_1 | e, \theta) = \frac{\alpha_1(1) * \beta_1(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.05 \times 0.0009673975}{0.0001897154} = 0.2549602$$
$$P(X_1 = S_2 | e, \theta) = 1 - P(X_1 = S_1 | e, \theta) = 0.7450398$$
$$P(X_2 = S_1 | e, \theta) = \frac{\alpha_2(1) * \beta_2(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.038 \times 0.00290725}{0.0001897154} = 0.5823223$$
$$P(X_2 = S_2 | e, \theta) = 1 - P(X_2 = S_1 | e, \theta) = 0.4176777$$
$$P(X_3 = S_1 | e, \theta) = \frac{\alpha_3(1) * \beta_3(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.003125 \times 0.012535}{0.0001897154} = 0.2064771$$
$$P(X_3 = S_2 | e, \theta) = 1 - P(X_3 = S_1 | e, \theta) = 0.7935229$$
$$P(X_4 = S_1 | e, \theta) = \frac{\alpha_4(1) * \beta_4(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.00048575 \times 0.0925}{0.0001897154} = 0.2368383$$
$$P(X_4 = S_2 | e, \theta) = 1 - P(X_4 = S_1 | e, \theta) = 0.7631617$$
$$P(X_5 = S_1 | e, \theta) = \frac{\alpha_5(1) * \beta_5(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.00048005 \times 0.31}{0.0001897154} = 0.7844145$$
$$P(X_5 = S_2 | e, \theta) = 1 - P(X_5 = S_1 | e, \theta) = 0.2155855$$
$$P(X_6 = S_1 | e, \theta) = \frac{\alpha_6(1) * \beta_6(1)}{\alpha_6(1) + \alpha_6(2)} = \frac{0.0001602455 \times 1}{0.0001897154} = 0.8446627$$
$$P(X_6 = S_2 | e, \theta) = 1 - P(X_6 = S_1 | e, \theta) = 0.1553373$$

2.3 Viterbi Algorithm

For solving this problem I would take the log of all values to make the calculations easier and add the log of probabilities instead of multiplying. The new values of the probabilities after taking log2(p) where p are the probabilities are:
Log base 2 of Transition probabilities:

$$P(S1|S1) = -0.5145, P(S2|S1) = -1.7369$$
$$P(S1|S2) = -1.7369, P(S2|S2) = -0.5145$$

Log base 2 of Emission probabilities:

$$P(A|S1) = -1.3219, P(C|S1) = -3.3219, P(G|S1) = -1.3219, P(T|S1) = -3.3219$$
$$P(A|S2) = -3.3219, P(C|S2) = -1.3219, P(G|S2) = -3.3219, P(T|S2) = -1.3219$$

Log base 2 of Initial state distribution:

$$\pi_1 = -1, \pi_2 = -1$$

Since this is a dynamic programming algorithm, we will store the probabilities of most probable paths and also store a pointer for which state that probability was calculated from so that we can backtrack the path later. The probability of the most probable path ending in state k with observation "i" at position t is:

$$p_k(i,t) = e_k(i) \max_j (p_j(o,t-1).p(jk))$$

Substituting in the above formula, we get the following values:

$$
\begin{aligned}
p_{S_1}(C,1) &= -1 - 3.3219 = -4.3219 \\
p_{S_2}(C,1) &= -1 - 1.3219 = -2.3219 \\
p_{S_1}(G,2) &= -1.3219 + \max((p_{S_1}(C,1) + p_{11}), (p_{S_2}(C,1) + p_{21})) \\
&= -1.3219 + \max(-4.3219 - 0.5145, -2.3219 - 1.7369) \\
&= -5.3807(\text{ came from } \mathbf{S2}) \\
p_{S_2}(G,2) &= -3.3219 + \max((p_{S_1}(C,1) + p_{12}), (p_{S_2}(C,1) + p_{22})) \\
&= -3.3219 + \max(-4.3219 - 1.7369, -2.3219 - 0.5145) \\
&= -6.1583(\text{ came from S2}) \\
p_{S_1}(T,3) &= -3.3219 + \max((p_{S_1}(G,2) + p_{11}), (p_{S_2}(G,2) + p_{21})) \\
&= -3.3219 + \max(-5.3807 - 0.5145, -6.1583 - 1.7369) \\
&= -9.2171(\text{ came from S1}) \\
p_{S_2}(T,3) &= -1.3219 + \max((p_{S_1}(G,2) + p_{12}), (p_{S_2}(G,2) + p_{22})) \\
&= -1.3219 + \max(-5.3807 - 1.7369, -6.1583 - 0.5145) \\
&= -7.9947(\text{ came from } \mathbf{S2}) \\
p_{S_1}(C,4) &= -3.3219 + \max((p_{S_1}(T,3) + p_{11}), (p_{S_2}(T,3) + p_{21})) \\
&= -3.3219 + \max(-9.2171 - 0.5145, -7.9947 - 1.7369) \\
&= -13.0535(\text{ can come from either S1 or S2}) \\
p_{S_2}(C,4) &= -1.3219 + \max((p_{S_1}(T,3) + p_{12}), (p_{S_2}(T,3) + p_{22})) \\
&= -1.3219 + \max(-9.2171 - 1.7369, -7.9947 - 0.5145) \\
&= -9.8311(\text{ came from } \mathbf{S2}) \\
p_{S_1}(A,5) &= -1.3219 + \max((p_{S_1}(C,4) + p_{11}), (p_{S_2}(C,4) + p_{21})) \\
&= -1.3219 + \max(-13.0535 - 0.5145, -9.8311 - 1.7369) \\
&= -12.8899(\text{ came from } \mathbf{S2}) \\
p_{S_2}(A,5) &= -3.3219 + \max((p_{S_1}(C,4) + p_{12}), (p_{S_2}(C,4) + p_{22})) \\
&= -3.3219 + \max(-13.0535 - 1.7369, -9.8311 - 0.5145) \\
&= -13.6675(\text{ came from S2}) \\
p_{S_1}(G,6) &= -1.3219 + \max((p_{S_1}(A,5) + p_{11}), (p_{S_2}(A,5) + p_{21})) \\
&= -1.3219 + \max(-12.8899 - 0.5145, -13.6675 - 1.7369) \\
&= -14.7263(\text{ came from } \mathbf{S1}) \\
p_{S_2}(G,6) &= -3.3219 + \max((p_{S_1}(A,5) + p_{12}), (p_{S_2}(A,5) + p_{22})) \\
&= -3.3219 + \max(-12.8899 - 1.7369, -13.6675 - 0.5145) \\
&= -17.5039(\text{ came from S2})
\end{aligned}
$$

We also have to check for final step so we will just check the max term to see which state the is more likely to be in the path

$$\begin{aligned}
p_{S_1}(o_7,7) &= \max((p_{S_1}(G,6)+p_{11}),(p_{S_2}(G,6)+p_{21})) \\
&= \max(-14.7263-0.5145,-17.5039-1.7369) \\
&= -15.2408(\text{ came from } \mathbf{S1}) \\
p_{S_2}(o_7,7) &= \max((p_{S_1}(G,6)+p_{12}),(p_{S_2}(G,6)+p_{22})) \\
&= \max(-14.7263-1.7369,-17.5039-0.5145) \\
&= -16.4632(\text{ came from S1})
\end{aligned}$$

We see that the sequence of most likely states estimated independently is $S_2S_1S_2S_2S_1S_1$ whereas the most likely path as calculated above is $S_2S_2S_2S_2S_1S_1$ and is indicated in bold in the above calculations. Since we took the log of probabilities, therefore the overall probability of the most likely sequence is $2^{(\sum p)}$ where p is the probability of state in the likely path.

2.4 Prediction

We would calculate all possible combinations and see which combinations gives maximum probability. We will calculate the forward probabilities as below. Let the observation at t=7 be o_7

$$\begin{aligned}
\alpha_7(1) &= (\alpha_6(1)a_{11}+\alpha_6(2)a_{21})b_1(o_7) \\
\alpha_7(2) &= (\alpha_6(1)a_{12}+\alpha_6(2)a_{22})b_2(o_7)
\end{aligned}$$

Since the emission probabilities of A and G are same given the state and probabilities of C and T are same given the state, therefore we can calculate for only A and C and use that for G and T

$$\begin{aligned}
b_1(A) = b_1(G) &\text{ and } b_2(A) = b_2(G) \\
b_1(C) = b_1(T) &\text{ and } b_2(C) = b_2(T)
\end{aligned}$$

Let $o_7 = A$

$$\begin{aligned}
\alpha_7(1) &= (\alpha_6(1)a_{11}+\alpha_6(2)a_{21})b_1(A) \\
&= (0.0001602455\times 0.7+0.00002946987\times 0.3)\times 0.4 \\
&= 0.000484051244 \\
\alpha_7(2) &= (\alpha_6(1)a_{12}+\alpha_6(2)a_{22})b_2(A) \\
&= (0.0001602455\times 0.3+0.00002946987\times 0.7)\times 0.1 \\
&= 0.0000068702559
\end{aligned}$$

Therefore P(A) = P(G) = $P(A|S_1) + P(A|S_2)$ = 0.0000552753803
Let $o_7 = C$

$$\begin{aligned}\alpha_7(1) &= (\alpha_6(1)a_{11} + \alpha_6(2)a_{21})b_1(C)\\ &= (0.0001602455 \times 0.7 + 0.00002946987 \times 0.3) \times 0.1\\ &= 0.0000121012811\\ \alpha_7(2) &= (\alpha_6(1)a_{12} + \alpha_6(2)a_{22})b_2(C)\\ &= (0.0001602455 \times 0.3 + 0.00002946987 \times 0.7) \times 0.4\\ &= 0.0000274810236\end{aligned}$$

Therefore P(C) = P(T) = $P(C|S_1) + P(C|S_2)$ = 0.0000395823047
We have to divide both the terms by $P(e|\theta)$ which we calculated in the first part of this question. But since we have to divide both of them we can avoid it and make a relative prediction. We see that the likelihood of seeing A or G is more than C or T and therefore the most likely emitted symbol in the next iteration would be A or G.

3 Programming

3.0.1 Eigenfaces

3.0.2 PCA Linear SVM

Optimal C	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
d = 20	0.1000	0.1000	0.0000	$1.0e^{11}$	$1.0e^{11}$
d = 50	0.1000	0.0100	0.0000	$1.0e^{11}$	$1.0e^{11}$
d = 100	0.1000	0.1000	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$
d = 200	0.0100	0.1000	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$

3.0.3 PCA RBF kernel SVM

Optimal C	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
d = 20	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$
d = 50	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$
d = 100	$1.0e^{10}$	$1.0e^{10}$	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{11}$
d = 200	$1.0e^{11}$	$1.0e^{11}$	$1.0e^{10}$	$1.0e^{11}$	$1.0e^{11}$

Optimal gamma	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
d = 20	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$
d = 50	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$
d = 100	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$
d = 200	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$	$1.0e^{-10}$

Figure 3.1: First eigenface

Figure 3.2: Second eigenface

Figure 3.3: Third eigenface

Figure 3.4: Fourth eigenface

Figure 3.5: Fifth eigenface

d	Linear SVM	RBF kernel
d = 20	86.5163%	83.3283%
d = 50	90.1529%	86.8033 %
d = 100	90.9311%	85.7594 %
d = 200	90.8571%	86.0100%

3.1 HIDDEN MARKOV MODEL

The estimates from running the model for 500 iterations are:

$$A = \begin{pmatrix} 0.9189 & 0.0811 \\ 0.0720 & 0.9280 \end{pmatrix}$$

$$E = \begin{pmatrix} 0.0957 & 0.4325 & 0.3943 & 0.0775 \\ 0.3860 & 0.1033 & 0.1153 & 0.3953 \end{pmatrix}$$

The estimates from hmm package are :

$$A_{hmm} = \begin{pmatrix} 0.9249 & 0.0751 \\ 0.0844 & 0.9156 \end{pmatrix}$$

$$E_{hmm} = \begin{pmatrix} 0.0891 & 0.4386 & 0.4003 & 0.0720 \\ 0.3850 & 0.1048 & 0.1165 & 0.3937 \end{pmatrix}$$

We see that these values are similar.