CSCI 567 : ASSIGNMENT- 5

NAME: NEEL SHAH

USC ID: 5417391181

Email id: neelshah@usc.edu

SOLUTION 1

PCA finds the first priniciple component vector $\vec{v}$ by minimizing the following objective function :

$$J(\vec{v}) = \sum_{i=1}^{n} ||\vec{x_i} - (\vec{v}^T\vec{x_i})\vec{v}||^2$$

$$\vec{v} \ is \ a \ unit \ vector \ which \ means \ ||\vec{v}|| = 1$$

$$R(\vec{v}) = \sum_{i=1}^{n} (\vec{v}^T\vec{x_i})^2 = \vec{v}^T X X^T \vec{v}$$

Expanding the expression $J(\vec{v})$:

$$\begin{aligned} J(\vec{v}) &= \sum_{i=1}^{n} ||\vec{x_i}||^2 + \sum_{i=1}^{n} ||(\vec{v}^T\vec{x_i})\vec{v}||^2 - \sum_{i=1}^{n} 2\vec{v}^T\vec{x_i}^T\vec{x_i}\vec{v} \\ &= \sum_{i=1}^{n} ||\vec{x_i}||^2 + \sum_{i=1}^{n} ||\vec{v}\vec{v}^T\vec{x_i}||^2 - \sum_{i=1}^{n} 2\vec{v}^T\vec{x_i}^T\vec{x_i}\vec{v} \\ &since \ \vec{v}\vec{v}^T = 1 \\ &= -\sum_{i=1}^{n} 2\vec{v}^T\vec{x_i}^T\vec{x_i}\vec{v} \\ &= -R(\vec{v}) \end{aligned}$$

Therefore, minimizing $J(\vec{v})$ and maximizing $R(\vec{v})$ are equivalent.

SOLUTION 2

1 FORWARD ALGORITHM

$$\alpha_1(X_1) = P(X_1)P(Y_1|X_1)$$
$$\alpha_1(S_1) = P(X_1 = S_1)P(y = 0|X_1 = S_1)$$
$$= 0.6 * 0.7 = 0.42$$
$$\alpha_1(S_2) = P(X_1 = S_2)P(y = 0|X_1 = S_2)$$
$$= 0.4 * 0.8 = 0.32$$

$Consider$

$$\alpha_t(X_k) = P(Y_t|X_k)\sum_i \alpha_{t-1}(S_i) * P(X_k|X_i)$$
$$\alpha_2(S_1) = P(y = 1|S_1)\sum_i \alpha_1(S_i) * P(S_1|S_i)$$
$$\alpha_2(S_2) = P(y = 1|S_2)\sum_i \alpha_1(S_i) * P(S_2|S_i)$$
$$i\ =\ 1\ ,\ 2$$
$$\alpha_2(S_1) = 0.3[(0.42 * 0.9) + (0.32 * 0.2)] = 0.1326$$
$$\alpha_2(S_2) = 0.2[(0.42 * 0.1) + (0.32 * 0.8)] = 0.0596$$

$$\alpha_3(S_1) = 0.7[(0.1326 * 0.9) + (0.0596 * 0.2)] = 0.091882$$
$$\alpha_3(S_2) = 0.8[(0.1326 * 0.1) + (0.0596 * 0.8)] = 0.048752$$

$The\ probability\ of\ observing\ the\ sequence\ is:$

$$P(Y_1 = 0, Y_2 = 1, Y_0 = 0) = \sum_i \alpha_3(S_i) = 0.91882 + 0.048752 = 0.140634$$

2 BACKWARD ALGORITHM

$$\begin{aligned}
\beta_3(S_1) &= 1 \\
\beta_3(S_2) &= 1
\end{aligned}$$

$$\begin{aligned}
\beta_t(X_k) &= \sum_i \beta_{t+1}(S_i) P(Y_{t+1}|S_i) P(S_i|S_k) \\
&t = 2, 1 \\
&i = 1, 2 \\
\beta_2(S_1) &= \sum_i \beta_3(S_i) P(Y_3|S_i) P(S_i|S_1) \\
&= (1) * (0.7) * (0.9) + (1) * (0.8) * (0.1) = 0.71 \\
\beta_2(S_2) &= \sum_i \beta_3(S_i) P(Y_3|S_i) P(S_i|S_2) \\
&= (1) * (0.7) * (0.2) + (1) * (0.8) * (0.8) = 0.78
\end{aligned}$$

$$\begin{aligned}
\beta_1(S_1) &= \sum_i \beta_2(S_i) P(Y_2|S_i) P(S_i|S_1) \\
&= (0.71) * (0.3) * (0.9) + (0.78) * (0.2) * (0.1) = 0.2073 \\
\beta_1(S_2) &= \sum_i \beta_2(S_i) P(Y_2|S_i) P(S_i|S_2) \\
&= (0.71) * (0.3) * (0.2) + (0.78) * (0.2) * (0.8) = 0.1674 \\
&The\ probability\ of\ observing\ the\ sequence\ is: \\
P(Y_1 = 0, Y_2 = 1, Y_0 = 0) &= \sum_i \alpha_i(S_i) * \beta_i(S_i) \\
&= (0.42) * (0.2073) + (0.32) * (0.1674) \\
&= 0.1406
\end{aligned}$$

3 Yes, the results from forward and backward algorithm agree.

4

The α and *beta* are computed above. Using the forward-backward algorithm, the most likely setting for each state is computed as follows :

$$
\begin{aligned}
\gamma_1(S_1) &= \frac{\alpha_1(S_1)\beta_1(S_1)}{\alpha_1(S_1)\beta_1(S_1) + \alpha_1(S_2)\beta_1(S_2)} \\
&= \frac{(0.42)(0.2073)}{(0.42)(0.2073) + (0.32)(0.1674)} = 0.619096 \\
\gamma_1(S_2) &= \frac{\alpha_1(S_2)\beta_1(S_2)}{\alpha_1(S_1)\beta_1(S_1) + \alpha_1(S_2)\beta_1(S_2)} \\
&= \frac{(0.32)(0.1674)}{(0.42)(0.2073) + (0.32)(0.1674)} = 0.3809 \\
\gamma_2(S_1) &= \frac{\alpha_2(S_1)\beta_2(S_1)}{\alpha_2(S_1)\beta_2(S_1) + \alpha_2(S_2)\beta_2(S_2)} \\
&= \frac{(0.1326)(0.71)}{(0.1326)(0.71) + (0.0596)(0.78)} = 0.6694 \\
\gamma_2(S_2) &= \frac{\alpha_2(S_2)\beta_2(S_2)}{\alpha_2(S_1)\beta_2(S_1) + \alpha_2(S_2)\beta_2(S_2)} \\
&= \frac{(0.0596)(0.78)}{(0.1326)(0.71) + (0.0596)(0.78)} = 0.3306 \\
\gamma_3(S_1) &= \frac{\alpha_3(S_1)\beta_3(S_1)}{\alpha_3(S_1)\beta_3(S_1) + \alpha_3(S_2)\beta_3(S_2)} \\
&= \frac{(0.091882)(1)}{(0.091882)(1) + (0.048752)(1)} = 0.6533 \\
\gamma_3(S_2) &= \frac{\alpha_3(S_2)\beta_3(S_2)}{\alpha_3(S_1)\beta_3(S_1) + \alpha_3(S_2)\beta_3(S_2)} \\
&= \frac{(0.048752)(1)}{(0.091882)(1) + (0.048752)(1)} = 0.3467
\end{aligned}
$$

5

The Viterbi algorithm predicts that the most likely sequence of states is A, A, A. The relevant computations are:

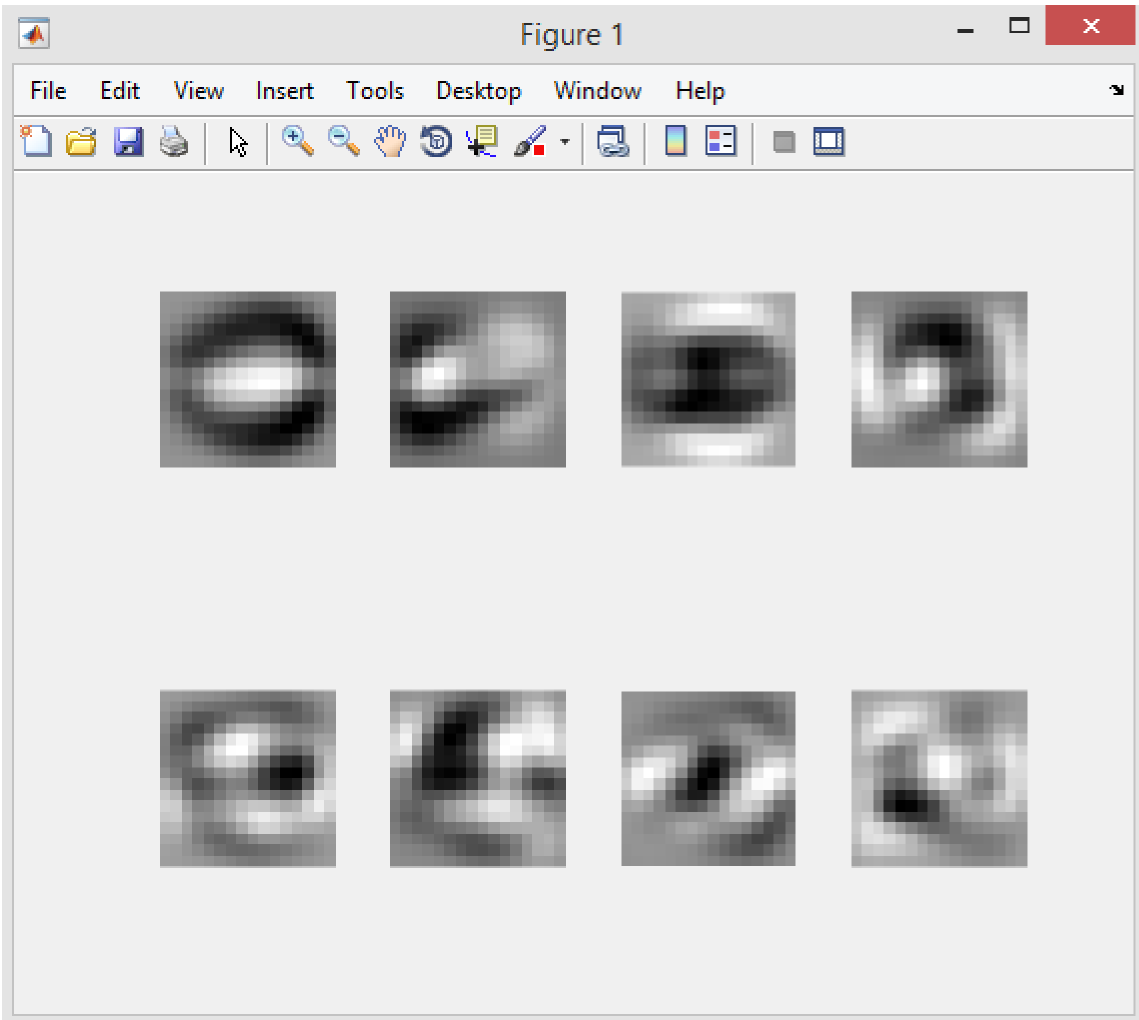
$$
\begin{aligned}
&Calculating\ initial\ V_1(S_1)\ V_1(S_2)\\
V_1(S_1) &= P(y=0|S_1)P(S_1)\\
&= 0.7 * 0.6 = 0.42\\
V_1(S_2) &= P(y=0|S_2)P(S_2)\\
&= 0.8 * 0.4 = 0.32\\
Using\ V_t(k) &= P(O_t|S_t=k) * argmax_i(P(S_t=k|S_{t-1}=i))V_{t-1}(k)\\
i\ &= 1,2\ ;\ t\ = 2,3\ k = S_1, S_2\\
V_2(S_1) &= P(y=1|S_1) * argmax[P(S_1|S_1)V_1(S_1), P(S_1|S_2)V_1(S_2)]\\
&= 0.3 * argmax[0.9 * 0.42, 0.2 * 0.32]\\
&= 0.3 * 0.9 * 0.42\\
&= 0.1134\\
V_2(S_2) &= P(y=1|S_2) * argmax[P(S_2|S_1)V_1(S_1), P(S_2|S_2)V_1(S_2)]\\
&= 0.2 * argmax[0.1 * 0.42, 0.8 * 0.32]\\
&= 0.2 * 0.8 * 0.32\\
&= 0.0512\\
\\
V_3(S_1) &= P(y=0|S_1) * argmax[P(S_1|S_1)V_2(S_1), P(S_1|S_2)V_2(S_2)]\\
&= 0.7 * argmax[0.9 * 0.1134, 0.2 * 0.0512]\\
&= 0.7 * 0.9 * 0.1134\\
&= 0.071442\\
V_3(S_2) &= P(y=0|S_2) * argmax[P(S_2|S_1)V_2(S_1), P(S_2|S_2)V_2(S_2)]\\
&= 0.8 * argmax[0.1 * 0.1134, 0.8 * 0.0512]\\
&= 0.8 * 0.8 * 0.0512\\
&= 0.0328
\end{aligned}
$$

Therefore, the most likely sequence of states using Viterbi Algorithm is S_1, S_1, S_1

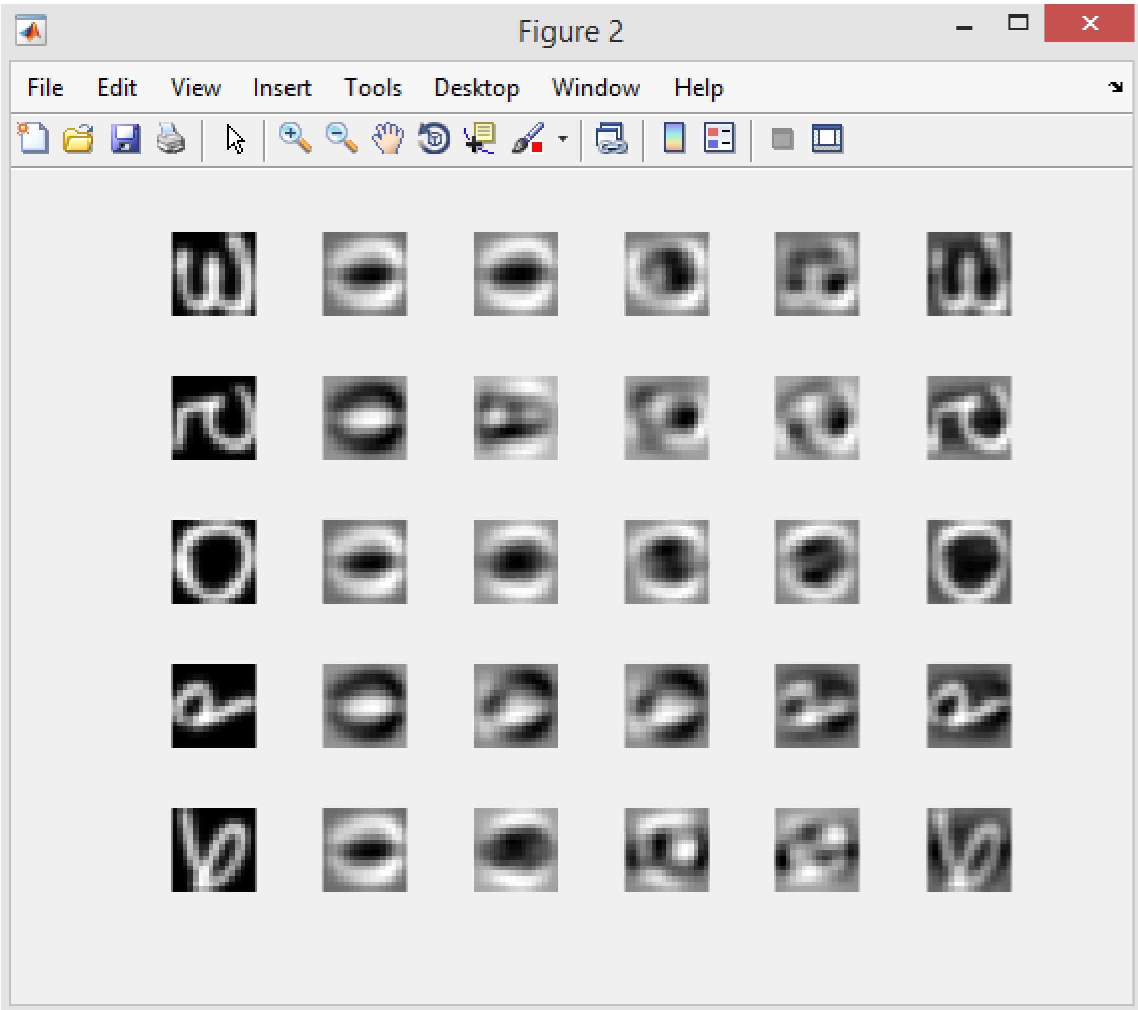
Q.2.e

	T=1	T=2	T=3
S1	0.42	0.1134	0.071442
S2	0.32	0.0512	0.0328

Q.3 b. Top 8 eigendigits:



c.



d.

Training data:

K	**Accuracy**	**Time (seconds)**
1	60.1527	0.12687
3	86.6960	0.01306
5	95.3753	0.00946
15	97.7952	0.00802
100	97.5694	0.00837

Test data:

K	**Accuracy**	**Time (seconds)**
1	54.6364	0.00740
3	79.5000	0.00584
5	87.1818	0.00462
15	89.2273	0.00525
100	88.9545	0.00397