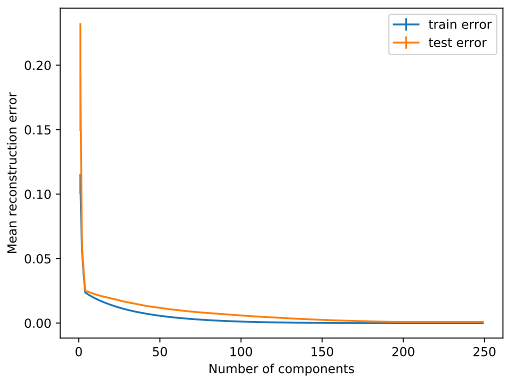
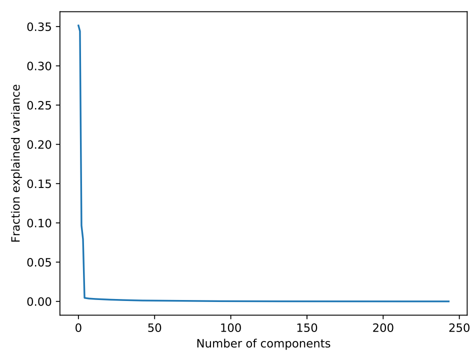
COMP 652: ASSIGNMENT 3

CARLOS G. OLIVER (260425853)

1. Q1: PCA



(A)



(B)

FIGURE 1

2. Q2: SPECTRAL METHODS FOR WEIGHTED AUTOMATA

2.1. **(a).** We use a property of Hankel matrices which is: if $\text{rank}(H_f) = n \Rightarrow$ there exists a weighted finite automaton $\mathcal{A}$ with n states such that $g = g_{\mathcal{A}}$ for some function $g : \Sigma^* \rightarrow \mathbb{R}$ where Σ^* is the set of all strings that can be generated from alphabet Σ. In this example, the function g counts the number of 1s in a string generated from $\Sigma = \{0, 1\}$ the Hankel matrix H_g would take the form:

Date: April 12, 2017.

$$H_g = \begin{array}{c} \\ \lambda \\ 0 \\ 1 \\ 11 \\ 111 \\ 110 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} \lambda & 0 & 1 & 11 & \dots \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 2 & \dots \\ 0 & 0 & 1 & 2 & \\ 1 & 1 & 2 & 3 & \\ 2 & 2 & 3 & 4 & \\ 3 & 3 & 4 & 5 & \\ 2 & 2 & 3 & 4 & \\ \vdots & & & & \ddots \end{bmatrix} \end{array}$$

The above matrix is arranged so all possible strings composed only of 1s, $Q = \{1\}^N \quad \forall N \in \mathbb{Z}^+$ precede all other strings $(\Sigma^* - Q)$ with the exception of $\{0\}^1$ and λ. If we order prefixes and suffixes in S in increasing N we can easily see that any row m in this sub block H_g^1 of H_g is a linear combination in the form $H_g^1(m,\cdot) = 2H_g^1(m,\cdot) - H_g^1(m-1,\cdot)$. The remaining entries in H_g are formed by prefixes and suffixes containing 0s which do not contribute to the evaluation of g and therefore can also be obtained from rows or columns in H^1. Therefore the rank of H_g is 2 where only the λ or $\{0\}^1$ and the $\{1\}^1$ contribute to the row rank. The same can be shown for the column rank.

2.2. **(b).** If f is a probability distribution over Σ^* then we have $\sum_{s\in\Sigma^*} f(s) = 1$ gives a probability for every string in Σ^*. Then $f_{sub}(w)$ is the probability that the string w occurs in any word generated by Σ^* since u and v represent all possible prefixes and suffixes to the word w.

2.3. **(c).** Let $\Sigma_{w\in\Sigma^*} A_w \equiv S$.

$$\begin{aligned} S = (I - A_0 - A_1)^{-1} = (I - A_0 - A_1)S &= I \\ (I - A_0 - A_1)SS^{-1} &= S^{-1} \\ (I - A_0 - A_1)^{-1} &= S \end{aligned} \tag{1}$$

Now we use this identity to compute the sum of the function f over all words as $\Sigma_{w\in\Sigma^*} f(w)$. We can express the function over a string as

$$f(w) = \alpha_0^T A w \alpha_\infty \tag{2}$$

So for all the words we have

$$\begin{aligned} \sum_{w\in\Sigma^*} f(w) &= \sum_{w\in\Sigma^*} \alpha_0^T A_w \alpha_\infty \\ &= \alpha_0^T \Big(\sum_{w\in\Sigma^*} A_w \Big) \alpha_\infty \\ &= \alpha_0^T (I - A_0 - A_1)^{-1} \alpha_\infty \end{aligned} \tag{3}$$

We can use this property to compute $f_{subsrting}(w)$ using an automaton by recognizing that the contribution of transition matrices from u and v to the sum can be reduced to the same identity as above as they sum over all of Σ^*.

$$\begin{aligned}
f_{substring}(w) &= \sum_{u\in\Sigma^*, v\in\Sigma^*} f(uwv) \\
&= \sum_{u\in\Sigma^*, v\in\Sigma^*} \alpha_0^T A_u A_w A_v \alpha_\infty \\
&= \alpha_0^T (I - A_0 - A_1)^{-1} A_w (I - A_0 - A_1)^{-1} \alpha_\infty
\end{aligned} \tag{4}$$

2.4. **(d).** Because the word uwv is in Σ^* and the training sample S is drawn from f also over Σ^*, having an automaton that represents this function would also give us a way to learn the function $f_{substring}$. From S we can construct the empirical Hankel matrix $\hat{H}_{f_{substring}}$ and recover the relevant sub-blocks. We can then perform SVD on H and solve for the necessary transition weights A_w and initial/final vecctors α_0, α_∞ from which we can produce the weighted automaton $\hat{A}$. In order to recover an estimate of f we would have to sample from the automaton over all $w \in \Sigma^*$ as this would compute the probability distribution over all words in Σ^*.

3. Q3: Method of moments and multiview model

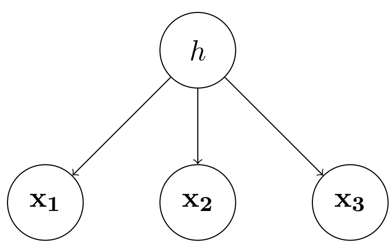


Figure 2. Graphical representation of conditional independence relation between random variables $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ and h.

Since the $\mathbf{x_t}$ are conditionally independent given h we can decompose the conditional expectation of the cross moments.

$$\begin{aligned}
\mathbb{E}[\mathbf{x_1} \otimes \mathbf{x_2} | h = j] &= \mathbb{E}[\mathbf{x_1} | h = j] \otimes \mathbb{E}[\mathbf{x_2} | h = j] && \text{by conditional independence} \\
&= \mu_{t=1,j} \otimes \mu_{t=2,j} \in \mathbb{R}^{d_1 \times d_2} && \text{by question statement} \\
&= \sum_{i=1}^{k} w_i \mu_{1,i} \otimes \mu_{2,i}
\end{aligned} \tag{5}$$

$$\mathbb{E}[\mathbf{x_1} \otimes \mathbf{x_2} \otimes \mathbf{x_3} | h = j] = \sum_{i=1}^{k} w_i \mu_{1,i} \otimes \mu_{2,i} \otimes \mu_{3,i} \in \mathbb{R}^{d_1 \times d_2 \times d_3} \tag{6}$$