

COMP 652: ASSIGNMENT 3

CARLOS G. OLIVER (260425853)

1. Q1: PCA

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:

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

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 .

Date: April 8, 2017.

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

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

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

$$(2) \quad f(w) = \alpha_0^T A_w \alpha_\infty$$

So for all the words we have

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

We can use this property to compute $f_{substring}(w)$ 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^* .

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

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 vectors 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^* .