CPSC340: Midterm

Friday 11th February 2011

Exercise 1. The bag-of-words representation of a paper is a vector with one dimension per word in the dictionary, recording how many times that word appeared in the paper.

(a) **[10 points]** Explain how you would combine the representations of all papers by a given author to get a single bag-of-words for that author.

(b) **[10 points]** Describe, in words, an algorithm not based on any probabilistic model for finding the three authors in a database whose work is most relevant to a given paper, and are not authors of the paper.

(c) **[10 points]** How could you use principal components analysis of bags of words to simplify and improve this system?

Exercise 2. The 1977 US Census reported the following features for each state of the USA:

Population	in thousands
Income	dollars per capita
Illiteracy	Percent of the adult population unable to read and write
Life Exp	Average years of life expectancy at birth
Murder	Number of murders and non-negligent manslaughters per 100,000 people
HS Grad	Percent of adults who were high-school graduates
Frost	Mean number of days per year with low temperatures below freezing
Area	In square miles

The summary statistics for these variables might be helpful.

Statistics	Min	Mean	Max
Population	365	4246	21198
Income	3098	4436	6315
Illiteracy	0.50	1.17	2.80
Life Exp	67.97	70.88	73.60
Murder	1.40	7.38	15.10
HS Grad	37.80	53.11	67.30
Frost	0.00	104.46	188.00
Area	1049	70736	566432

We will do two different principal component analyses of this data. For the first one, we do center the data but do not scale the features (`states.pca.1`) whereas for the second one we do center and normalize the features so that they are approximately on the same scale (`states.pca.2`) . We obtain the following results for `states.pca.1` (Table 1 and Figure 1) and `states.pca.2` (Table 2, Figure 2 and Figure 3).

	PC1	PC2
Population	1.18×10^{-03}	-1.00
Income	2.62×10^{-3}	-2.80×10^{-2}
Illiteracy	5.52×10^{-7}	-1.42×10^{-5}
Life Exp	-1.69×10^{-6}	1.93×10^{-5}
Murder	9.88×10^{-6}	-2.79×10^{-4}
HS Grad	3.16×10^{-5}	1.88×10^{-4}
Frost	3.61×10^{-5}	3.87×10^{-3}
Area	1.00	1.26×10^{-3}

Table 1: Projections of the features on to the first two principal components of `states.pca.1`

	PC1	PC2
Population	0.1260	0.4110
Income	−0.2990	0.5190
Illiteracy	0.4680	0.0530
Life Exp	−0.4120	−0.0817
Murder	0.4440	0.3070
HS Grad	−0.4250	0.2990
Frost	−0.3570	−0.1540
Area	−0.0334	0.5880

Table 2: Projections of the features on to the first two principal components of `states.pca.2`

(a) **[10 points]** Describe, in words, which features are most important under projection on to the first and second principal components for `states.pca.1` and `states.pca.2`.

(b) **[10 points]** Would you rather use `states.pca.1` or `states.pca.2` for further analysis? Pick one and explain your choice. (A choice with no or inadequate reasoning will get little or no credit.)

(c) **[10 points]** Figure 3 shows the states in their geographic locations, with the size of the label being proportional to the projection on to the first component (as per `states.pca.2`). What does this suggest about the interpretation of that component?

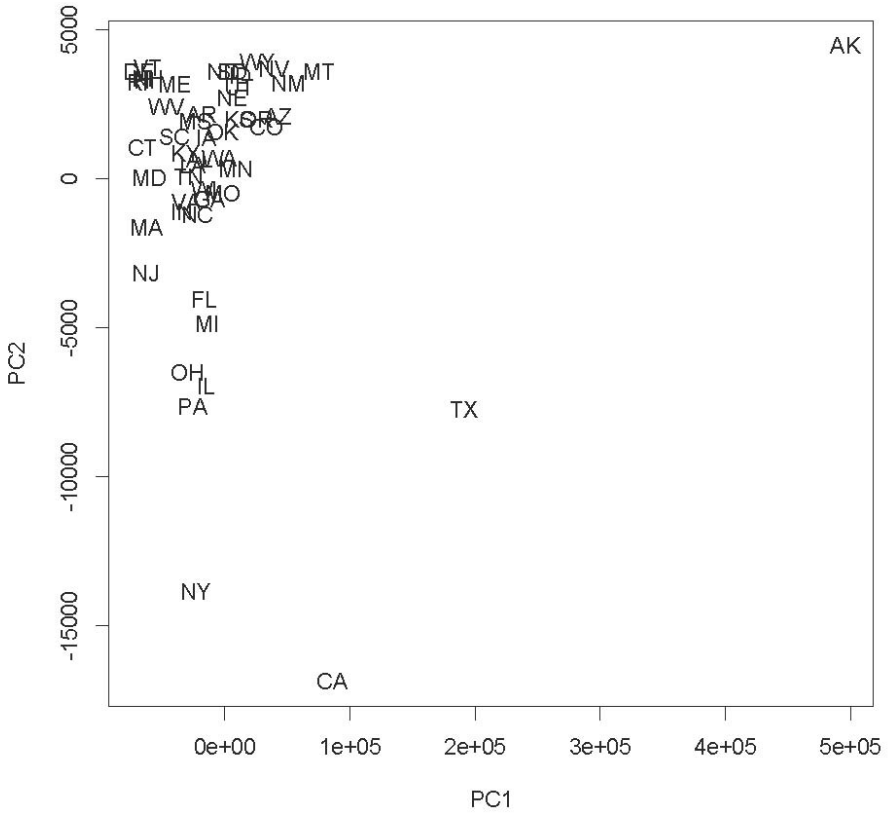

Figure 1: Projections of the states on to the first two principal components of `states.pca.1`.

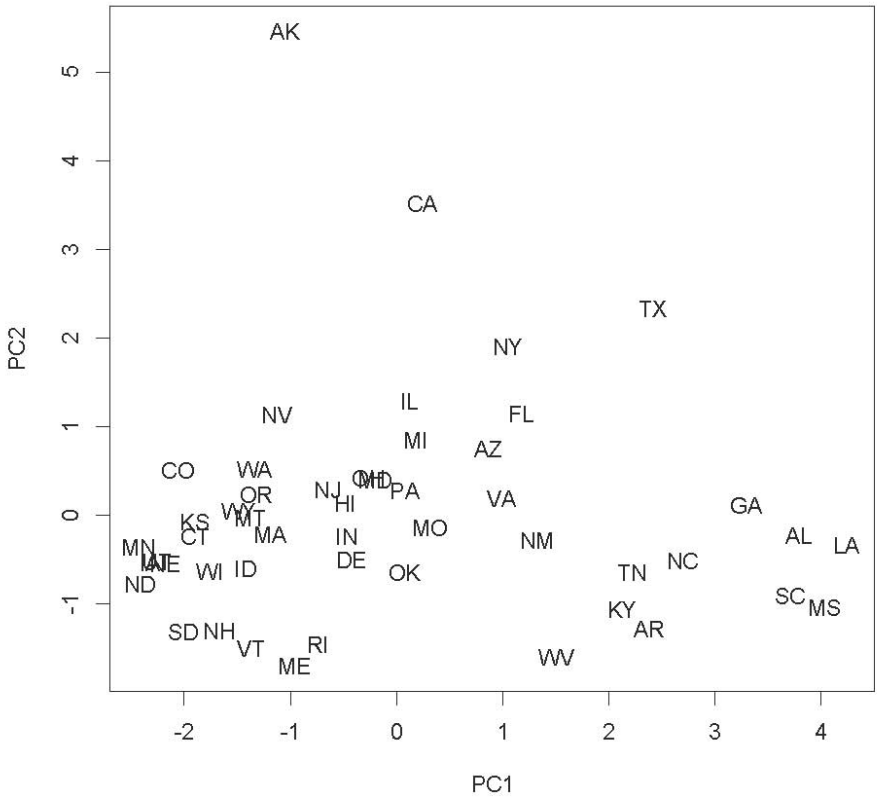

Figure 2: Projections of the states on to the first two principal components of `states.pca.2`.

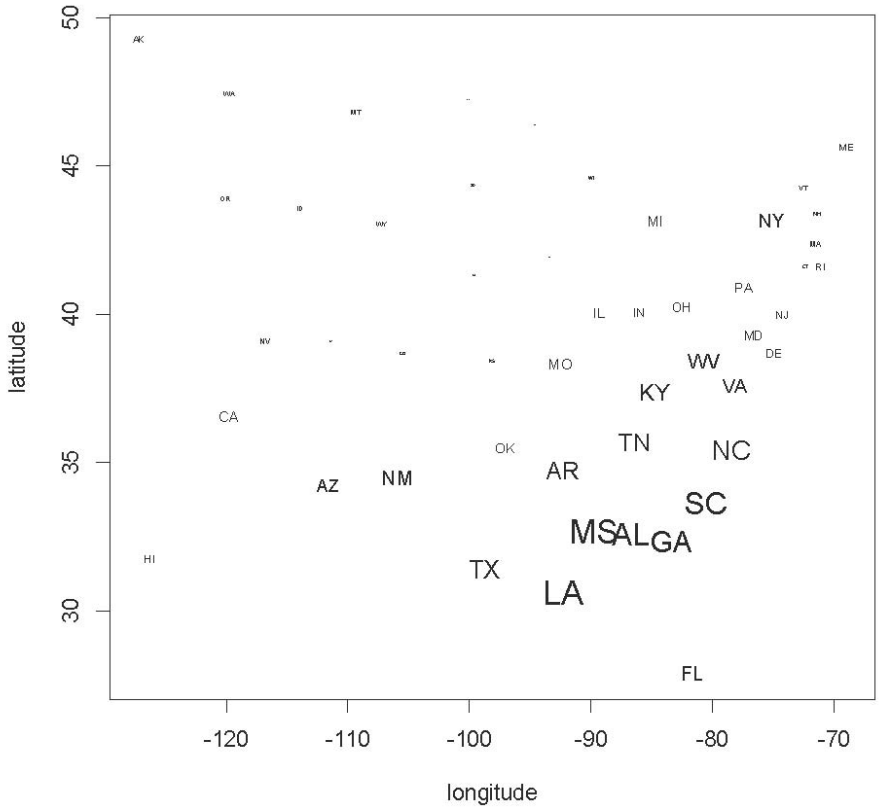

Figure 3: States in their geographic locations, with name size being proportional to the projection on to the first component of states.pca.2.

Exercise 3. Consider a Markov chain $\{X_k\}_{k\geq 0}$ such that

$$\Pr\left(X_k = j \middle| X_{k-1} = i\right) = T_{i,j}$$

where $i, j = 1, ..., n$ and define the associated $n \times n$ transition matrix T whose $(i,j)^{\text{th}}$ elements is $T_{i,j}$. Define also the column vector

$$\boldsymbol{\pi}_k = \left(\begin{array}{ccc} \Pr\left(X_k = 1\right) & \cdots & \Pr\left(X_k = n\right) \end{array} \right)^{\mathrm{T}}.$$

(a) **[10 points]** Prove that

$$\boldsymbol{\pi}_k = T^{\mathrm{T}}\ \boldsymbol{\pi}_{k-1}. \tag{1}$$

(b) **[15 points]** Consider the following transition matrices

$$T_1 = \left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right), T_2 = \left(\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \end{array} \right),$$

$$T_3 = \left(\begin{array}{cccc} \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{3}{5} & 0 & \frac{2}{5} \end{array} \right), T_4 = \left(\begin{array}{ccc} \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{array} \right).$$

Without doing any calculation, determine for these four cases whether the vector $\boldsymbol{\pi}_k = T^{\mathrm{T}}\ \boldsymbol{\pi}_{k-1}$ converges as $k \to \infty$ to a limit vector $\boldsymbol{\pi}$ that is *independent* of the initial distribution $\boldsymbol{\pi}_0$.

(c) **[15 points]** We assume here that $T_{i,j} > 0$ for all $i, j = 1, ..., n$. The power method is a general method to compute an eigenvector associated to the largest eigenvalue of a matrix. Starting from any initial vector $\mathbf{v}_0$, it proceeds as follows

$$\mathbf{w}_k = T^{\mathrm{T}}\ \mathbf{v}_{k-1},\ \mathbf{v}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \tag{2}$$

where $\boldsymbol{\pi} = T^{\mathrm{T}}\ \boldsymbol{\pi}$ is the unique column eigenvector of T^{T} with positive entries such that $\sum_{i=1}^{n} \boldsymbol{\pi}\left(i\right) = 1$.

Assume that we can diagonalize T^{T} using orthonormal eigenvectors $\boldsymbol{\pi}, \mathbf{u}_2, ..., \mathbf{u}_n$ associated to the eigenvalues $\lambda_1 = 1 > |\lambda_2| > \cdots > |\lambda_n|$ so that we can rewrite $\mathbf{v}_0 = a_1\boldsymbol{\pi} + \sum_{i=2}^{n} a_i\mathbf{u}_i$ where $a_1 = \mathbf{v}_0^{\mathrm{T}}\boldsymbol{\pi}$ and $a_i = \mathbf{v}_0^{\mathrm{T}}\mathbf{u}_i$ for $i = 2, ..., n$. Prove first that for any $k > 0$

$$\mathbf{v}_k = \frac{a_1\boldsymbol{\pi} + \sum_{i=2}^{n} \lambda_i^k a_i\mathbf{u}_i}{\left\|a_1\boldsymbol{\pi} + \sum_{i=2}^{n} \lambda_i^k a_i\mathbf{u}_i\right\|}.$$

Use this result to establish that if $\mathbf{v}_0 = \left(\mathbf{v}_0\left(1\right)\ \cdots \mathbf{v}_0\left(n\right)\right)$ is selected such that $\mathbf{v}_0\left(i\right) > 0$ for $i = 1, ..., n$ then for large k

$$\|\mathbf{v}_k - \ \boldsymbol{\pi}\| \approx \frac{|a_2|}{|a_1|}\left|\lambda_2\right|^k.$$