CPSC-340 Machine Learning and Data Mining 2004

Homework # 1

Due Friday, Sep 17 at 4pm.

NAME:______________________________

Signature:______________________________

STD. NUM: ______________________________

General guidelines for homeworks:

You are encouraged to discuss the problems with others in the class, but all write-ups are to be done on your own.

Homework grades will be based not only on getting the "correct answer," but also on good writing style and clear presentation of your solution. It is your responsibility to make sure that the graders can easily follow your line of reasoning.

Try every problem. Even if you can't solve the problem, you will receive partial credit for explaining why you got stuck on a promising line of attack. More importantly, you will get valuable feedback that will help you learn the material.

Please acknowledge the people with whom you discussed the problems and what sources you used to help you solve the problem (e.g. books from the library). This won't affect your grade but is important as academic honesty.

When dealing with Matlab exercises, please attach a printout with all your code and show your results clearly.

1. **(Learning Matlab)**

   (i) Write a MATLAB m file that plots the following 2D Gaussian density function:

   $$p(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

   for x and y in the range [-3,:0.1:3]. You will need to learn the commands `help`, `figure`, `clf`, `meshgrid` and `surf`. Your answer should consist of the `surf` plot and the code.

   (ii) Load the special MATLAB function called `peaks`. You can do this by typing `[x,y,z] = peaks;`. Use the commands `contour` and `clabel` to generate a countour plot of the function `peaks`. Repeat the exercise using the functions `contourf` and `contour3`. You should hand in your code and a <u>single</u> figure with 4 subplots (you'll need to learn the command `subplot`) consisting of the `surf` plot and the 3 countour plots.

   (iii) Now we'll try to visualise a function in 4D! In particular, use the command `slice` to generate a plot of the 3D contours of the following Gaussian:

   $$p(x, y, z) = \frac{1}{(2\pi)^{3/2}} e^{-\frac{1}{2}(x^2+y^2+z^2)}$$

   for x, y and z defined in the range [-3,:0.1:3].Use the commands `xlabel`, `ylabel` and `zlabel` to label the axes appropriately. Hand in the code and plot.

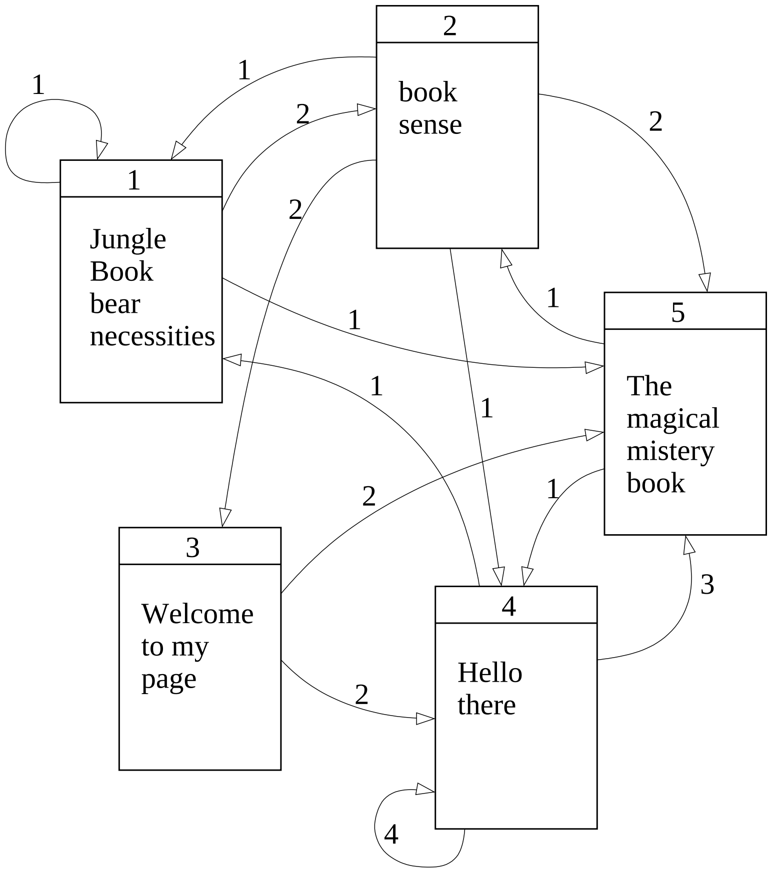
2. **(Eigenproblems)**

   (i) Suppose the matrix $\mathbf{A} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ has n linearly independent eigenvectors $\mathbf{x_1}, \ldots, \mathbf{x_n}$. Define the matrix $\mathbf{Q}$ having these vectors as columns, i.e. $\mathbf{Q} = [\mathbf{x_1} \quad \mathbf{x_2} \quad \ldots \quad \mathbf{x_n}]$. Let $\mathbf{D}$ be a diagonal matrix with the eigenvalues λ_i in the diagonal. Show that

   $$\mathbf{A} = \mathbf{QDQ^{-1}}$$

(ii) Using the previous decomposition of $\mathbf{A}$, what are the eigenvalues and eigenvectors of $\mathbf{A^2}$. Explain.

3. **(Google)** Given the following web graph:



(i) Construct the graph transition matrix. Write down the appropriately normalised version of this matrix and call it $\mathbf{G}$.

(ii) Using MATLAB, compute the rank vector using the PageRank algorithm described in class.

(iii) Write down a MATLAB script that plots the individual entries of the rank vector on the vertical axis and the number of iterations on the horizontal axis. This allows us to see how the algorithm is converging and determine how many iterations are necessary. Hand in the code and plot. Label the axes appropriately.

(iv) What are the eigenvalues of $\mathbf{G}^{\mathbf{1000}}$.

(v) For the query "book", what pages (in order) does this simple search engine return?