CS 485 Lecture 19 6 March 2006

Scribes: Jason Lui (jsl57) and Jeff Wong (jmw92)

Four generating functions from previous lecture

Deg: g_0, g_1

Size: h_0, h_1

$h_0'(1) = 1 + g_0'(1)h_1'(1)$ (I)

$h_1'(1) = 1 + g_1'(1)h_1'(1)$ (II)

From (II), we have

$h_1'(1) = 1 / [1 + g_1'(1)]$ (III)

Plugging (III) back to (I)

$h_0'(1) = 1 + g_0'(1) / [1 - g_1'(1)]$

Therefore, point at which the giant components appears when $g_1'(1) = 1$

We also know that:

$g_1'(1) = g_0''(1) / g_0'(1) = 1 \quad \rightarrow \quad g_0''(1) = g_0'(1)$

By definition:

$g_0(x) = \sum P_k x^k$

$g_0'(x) = \sum kP_k x^{k-1}$

$g_0''(x) = \sum k(k-1)P_k x^{k-2}$

Since $g_0''(1) = g_0'(1)$, let x = 1

$\sum kP_k = \sum k(k-1)P_k$

$\sum (k^2-2k)P_k = 0$

$\sum k(k-2)P_k = 0$

This is the end of random graph materials.

High Dimensional Data

10^6 papers

25,000 dimensional vector to represent the words used in the papers

This is a very large dimension, so we can randomly select a 300 dimensional subspace and project all the data points to such subspace. In fact, we can pick a subspace such that we minimize the error between data and its projection.

SVD: Singular Value Decomposition

Let **A** be the $10^6 \times 25000$ matrix of data from your set of papers. Any matrix **A** can be rewritten as $\mathbf{A}=\mathbf{U\Sigma V}^T$ where **U** and **V** are orthonormal and **Σ** is diagonalised:

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_N \end{pmatrix} \quad \text{with } \sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq \sigma_N$$

Create a new matrix $\mathbf{\Sigma}^{(k)}$ by keeping only the k largest σ's in $\mathbf{\Sigma}$:

$$\Sigma^{(k)} = \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & 0 & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & 0 & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

Then write $\mathbf{A}^{(k)} = \mathbf{U}\mathbf{\Sigma}^{(k)}\mathbf{V}^T$.

What is the error between $\mathbf{A}$ and $\mathbf{A}^{(k)}$?

Norms

- Frobenius norm: $|\mathbf{A}|_F^2 = \sum_{i,j} a_{ij}^2$ = sum of squares of all of the elements of $\mathbf{A}$.

  Let a_i be the ith column of $\mathbf{A}$. Then $(\mathbf{A}^T\mathbf{A})_{ij} = (a_i)^T a_j$.

  So $\mathrm{Tr}(\mathbf{A}^T\mathbf{A}) = \Sigma_i (a_i)^T a_i = \Sigma_i |a_i|^2$ = sum of squares of elements of A = $|\mathbf{A}|_F^2$

- 2-norm:

  For a vector $\mathbf{x} = (x_1, x_2, \ldots x_n)$, $|\mathbf{x}|_2 = \sqrt{x_1 + x_2 + x_3 + \ldots}$

  For a matrix, $|\mathbf{A}|_2 = \max_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|_2$

- It turns out that the 2-norm of A is the maximum eigenvalue of A, and the Frobenius norm squared is the sum of the squares of the eigenvalues. To see this, take the 2-norm and Frobenius norm of a diagonalised matrix and note the 2nd lemma below.

<u>Lemma</u>: $|\mathbf{AB}|_2 \leq |\mathbf{A}|_2 |\mathbf{B}|_2$

Pf: Let $\mathbf{y}$ be the value of $\mathbf{x}$ that maximises $|\mathbf{A}|_2 = \max_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|_2$.

Let $\mathbf{z} = \mathbf{B}\mathbf{y}$.

Then $|\mathbf{AB}|_2 = |\mathbf{A}\mathbf{z}|_2 = \left| \mathbf{A} \frac{\mathbf{z}}{|\mathbf{z}|} \right|_2 |\mathbf{z}|_2 \leq |\mathbf{A}|_2 |\mathbf{z}|_2 \leq |\mathbf{A}|_2 |\mathbf{B}|_2$ since $|\mathbf{z}|_2 \leq |\mathbf{B}|_2$

<u>Lemma</u>: Let Q be an orthonormal matrix. For all x, $|\mathbf{Q}\mathbf{x}|_2 = |\mathbf{x}|_2$.

Pf: $|\mathbf{Q}\mathbf{x}|_2^2 = \mathbf{x}^T\mathbf{Q}^T\mathbf{Q}\mathbf{X} = \mathbf{x}^T\mathbf{x} = |\mathbf{x}|_2^2$ because $\mathbf{Q}^T = \mathbf{Q}^{-1}$, so $\mathbf{Q}^T\mathbf{Q} = \mathbf{1}$.

<u>Lemma</u>: Let Q be an orthonormal matrix. For all $\mathbf{A}$, $|\mathbf{Q}\mathbf{A}|_2 = |\mathbf{A}|_2$.

Pf: This follows directly from the previous lemma.

These lemmas also hold for the Frobenius norm as well.

It is straightforward from this to show that for $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, $|\mathbf{A}| = |\mathbf{\Sigma}|$ for both norms.

<u>Thm</u>: $\left|\mathbf{A} - \mathbf{A}^{(k)}\right| \leq \min_{\mathrm{rank}(\mathbf{B}) \leq k} |\mathbf{A} - \mathbf{B}|$. In other words, $\mathbf{A}^{(k)}$ is the best rank-k approximation to $\mathbf{A}$ in terms of total error (but not necessarily error of a given element, so this is not always "best" in every application). Proof next lecture. This is easy to prove for the 2-norm but very difficult for the Frobenius norm.