Algorithm Foundations of Data Science and Engineering

Welcome Tutorial :-)

Tutorial 4-Suggested Answers

GAO Ming

DaSE @ ECNU

25 Mar., 2019

1. Given an input stream $< 4, 1, 3, 5, 1, 3, 2, 6, 7, 0, 9 >$ and hash functions in the form of $h(x) = (ax + b) \bmod 8$, where a and b are two arbitrary integers. If there are three following hash functions:

(1) $h(x) = (3x + 2) \bmod 8$;

(2) $h(x) = (7x + 5) \bmod 8$;

(3) $h(x) = (5x + 3) \bmod 8$;

Please address the following questions:

a. Find the frequency count of every item given by Count-Min sketch;

b. Analyze the accurate of counting result in *a*.;

c. If we try to find the (ε, δ)-approximations of the frequency count, how to modify the algorithm;

解：a：各项在三个哈希函数作用后的结果如下:

	4	1	3	5	2	6	7	0	9
h_1	6	5	3	1	0	4	7	2	5
h_2	1	4	2	0	3	7	6	5	4
h_3	7	0	2	4	5	1	6	3	0

其中项1和项9的哈希值冲突。

由count-min sketch算法，每项的频数计数和最终的估计值为：

	4	1	3	5	2	6	7	0	9
c_1	1	3	2	1	1	1	1	1	3
c_2	1	3	2	1	1	1	1	1	3
c_3	1	3	2	1	1	1	1	1	3
$\widehat{f}$	1	3	2	1	1	1	1	1	3
f	1	2	2	1	1	1	1	1	1

b:项1和项9高估了，其余项估计准确。

c:修改为$t=O(\log_2(\frac{1}{\delta}))$。因为

$$P(\widehat{f_a}-f_a\geq\varepsilon||f||_1)\leq P(\widehat{f_a}-f_a\geq\varepsilon||f_{-a}||_1)=P(minX_1,...,X_t\geq\varepsilon||f_{-a}||_1)$$

$$=\prod_{i=1}^{t}P(X_i\geq\varepsilon||f_{-a}||_1)\leq\frac{1}{2^t}<\delta$$

所以$t=O(\log_2(\frac{1}{\delta}))$

2. Let the largest and second largest eigenvalues of matrix A be 2 and 1.7, respectively. Is it possible to find the largest eigenvalue via using the power iteration approach? Please explain how fast the power method converges?

解：当最大特征值是一重时，能通过幂迭代法求出最大特征值。收敛速度是$|\frac{\lambda_2}{\lambda_1}|^k$

3. Given a matrix

$$A = \begin{pmatrix} 2 & 1 \\ 4 & 5 \end{pmatrix}$$

a. Compute the eigenvalues and eigenvectors of A;

b. Given a starting vector $v = (1,1)^T$, approximate the largest eigenvalue and eigenvector of A via using the power method.

解：a:令$|\lambda I - A| = (\lambda-1)(\lambda-6) = 0$,得到两个一重的特征值$\lambda_1 = 1, \lambda_2 = 6$。对于$\lambda_1 = 1, (A - \lambda_1 I)x = 0$

$$\left(\begin{bmatrix} 2 & 1 \\ 4 & 5 \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = \mathbf{0}.$$

特征向量为$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. 同理得到特征值$\lambda_2 = 6$对应的特征向量为$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

b: $v_0 = (1,1)^T, A = \begin{pmatrix} 2 & 1 \\ 4 & 5 \end{pmatrix},$

$$v_1^{'} = Av_0, v_1 = \frac{v_1^{'}}{\|v_1^{'}\|} = (0.3162, 0.9487)^T$$

$$v_2^{'} = Av_1, v_2 = \frac{v_2^{'}}{\|v_2^{'}\|} = (0.2545, 0.9671)^T$$

$$v_3^{'} = Av_2, v_3 = \frac{v_3^{'}}{\|v_3^{'}\|} = (0.2445, 0.9696)^T$$

$$v_4^{'} = Av_3, v_4 = \frac{v_4^{'}}{\|v_4^{'}\|} = (0.2429, 0.9701)^T$$

$$v_5^{'} = Av_4, v_5 = \frac{v_5^{'}}{\|v_5^{'}\|} = (0.2426, 0.9701)^T$$

$$v_6^{'} = Av_5, v_6 = \frac{v_6^{'}}{\|v_6^{'}\|} = (0.2425, 0.9701)^T$$

继续迭代满足$v_i - v_{i-1} \leq (0.0001, 0.0001)^T$,由幂迭代法可知，$Av_{max} = \lambda v_{max}$最大特征值约为6，最大特征向量为$(0.2425, 0.9701)^T$。

4. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be eigenvalues of matrix $A \in R^{n \times n}$. Prove that

a. The matrix $A - \sigma I$ has eigenvalue $\lambda_i - \sigma$ for $i = 1, 2, \cdots, n$;

b. The invertible matrix $(A - \sigma I)^{-1}$ has eigenvalue $(\lambda_i - \sigma)^{-1}$ for $i = 1, 2, \cdots, n$.

解：a:对λ_i ，$i = 1, 2, \cdots, n$，满足$Ax = \lambda_i x$。$(A - \sigma I)x = Ax - \sigma x = \lambda_i x - \sigma x = (\lambda_i - \sigma)x$，所以矩阵$A - \sigma I$的特征值是$\lambda_i - \sigma$。

b:$Ax = \lambda x$，两边同时乘A^{-1}，有$x = \lambda A^{-1} x$，$\frac{1}{\lambda} x = A^{-1} x$，所以$A^{-1}$ 的特征值为$\frac{1}{\lambda}$，所以$(A - \sigma I)^{-1}$的特征值是$(\lambda_i - \sigma)^{-1}$。