COMS 4771 Machine Learning (2018 Fall) Homework 2

Jing Qian - jq2282@columbia.edu

November 19, 2018

Problem 3

(Jing, Nov 14)

(i)

With the matrix A, binary vector x_i is hashed to b could be expressed as following:

$$\begin{aligned} b &= Ax_i \\ \begin{bmatrix} b^1 \\ b^2 \\ \cdots \\ b^p \end{bmatrix} &= \begin{bmatrix} A^1 \\ A^2 \\ \cdots \\ A^p \end{bmatrix} x_i \\ &= \begin{bmatrix} (\sum_{l=1}^{n} A^{1l}x_i^l) \bmod 2 \\ (\sum_{l=1}^{n} A^{2l}x_i^l) \bmod 2 \\ \cdots \\ (\sum_{l=1}^{n} A^{pl}x_i^l) \bmod 2 \end{bmatrix} \end{aligned} \tag{1}$$

Take the t-th element of b as an example: $b^t = ((\sum_{l=1}^{n} A^{tl}x_i^l) \bmod 2)$. Since $x^i \in \{0,1\}^n$, supposing there are n_i nonzero elements in x_i, then $\sum_{l=1}^{n} A^{tl}x_i^l$ is actually to randomly pick nonzero elements from the n_i nonzero elements in x_i. And $b^t = ((\sum_{l=1}^{n} A^{tl}x_i^l) \bmod 2)$ is in fact describing whether we pick even or odd number of nonzero elements from all the nonzero elements in x_i. Since A^t is picked uniformly at random, the probability of the t-th entry of b, $b^t = 0$ is $1/2$ and that of $b^t = 1$ is also $1/2$. And this is true for any t between 1 and p. Because different entries of matrix A are independent with each other, $\text{Prob}(x_i \to b) = \prod_{t=1}^{p} \text{Prob}(A^t x_i \to b^t) = \prod_{t=1}^{p} 1/2 = 1/2^p$.

(ii)

We could do this problem in two methods.

METHOD 1: From part (i), the probability of x_i hashing to any b is $1/2^p$, the probability of x_j hashing to any b is $1/2^p$. So the probability of x_j hashing to the same vector that x_i is hashing to is $1/2^p$.

METHOD 2: x_i and x_j hash to the same vector means: $Ax_i = Ax_j$

$$\begin{bmatrix} (\sum_{l=1}^{n} A^{1l}x_i^l) \bmod 2 \\ (\sum_{l=1}^{n} A^{2l}x_i^l) \bmod 2 \\ \cdots \\ (\sum_{l=1}^{n} A^{pl}x_i^l) \bmod 2 \end{bmatrix} = \begin{bmatrix} (\sum_{l=1}^{n} A^{1l}x_j^l) \bmod 2 \\ (\sum_{l=1}^{n} A^{2l}x_j^l) \bmod 2 \\ \cdots \\ (\sum_{l=1}^{n} A^{pl}x_j^l) \bmod 2 \end{bmatrix} \tag{2}$$

Then

$$\begin{bmatrix} (\sum_{l=1}^{n} A^{1l}[x_i^l - x_j^l]) \bmod 2 \\ (\sum_{l=1}^{n} A^{2l}[x_i^l - x_j^l]) \bmod 2 \\ \cdots \\ (\sum_{l=1}^{n} A^{pl}[x_i^l - x_j^l]) \bmod 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \tag{3}$$

which means $A(x_i - x_j) = 0$.

Supposing x_i and x_j have m differernt elements. Then any $A^t(x_i - x_j) = 0$ for $(1 \leq t \leq p)$ means to pick even number of elements from the m elements. This is the probability of a random variable X with binomial distribution Binomial(m, p_x) being even where $p_x = 1/2$ because this is a random and uniform case. So:

$$\begin{aligned} \text{Prob}(A^t(x_i - x_j) = 0) &= \text{Prob (X}_{m,p_x} \text{ is even)} \\ &= \frac{1}{2}(1 + (1 - 2p_x)^m) \\ &= \frac{1}{2}(1 + (1 - 2 \times \frac{1}{2})^m) \\ &= \frac{1}{2} \end{aligned} \tag{4}$$

Because diffferent elements of b are independent with each other, $\text{Prob}(A(x_i - x_j) = 0) = \prod_{t=1}^{p} \text{Prob}(A^t(x_j - x_j) = 0) = \prod_{t=1}^{p} 1/2 = 1/2^p$.

(iii)

The probability of no collisions among the x_i could be represented as following:

$$\begin{aligned}\text{Prob (no collisions)} &= 1 - \text{Prob (exist collisions)}\\ &\geq 1 - \sum_{1\leq i<j\leq m} \text{Prob}(x_i, x_j \text{ collide})\\ &= 1 - \sum_{1\leq i<j\leq m} 1/2^p\\ &= 1 - \binom{m}{2}\frac{1}{2^p}\\ &= 1 - \frac{m(m-1)}{2}\frac{1}{2^p}\\ &\geq 1 - \frac{m^2}{2}\frac{1}{2^p}\end{aligned} \tag{5}$$

If $p \geq 2\log_2 m$,

$$\begin{aligned}\text{Prob (no collisions)} &\geq 1 - \frac{m^2}{2}\frac{1}{2^p}\\ &\geq 1 - \frac{m^2}{2}\frac{1}{m^2}\\ &= 1 - 1/2\\ &= 1/2\end{aligned} \tag{6}$$

So if $p \geq 2\log_2 m$, there are no collisions among the x_i with probability at least 1/2.