COMS 4771 Machine Learning (2018 Fall)
Homework 2

Jing Qian - jq2282@columbia.edu

November 14, 2018

Problem 3

(Jing, Nov 14)

(i)

Since $b \in \{0,1\}^p$, there are 2^p possible bs. Since the entries of A are picked uniformly at random, the probability of x_i hashes to any b is equal and hence $1/2^p$. what's the point of the hint?

(ii)

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 $1/2^p$.

(iii)

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

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

If $p \geq 2 \log_2 m$,

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

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