Theoretical Computer Science III Term 2 Summative

wtxd25

March 16, 2020

1 Advanced Algorithms

Question 1

Let us partition $\mathscr{U}$ into the sets $\mathscr{U}_i = \{x : x \in \mathscr{U} \wedge h(x) = i\}$. It can be seen that the union of these subsets is the original set $\mathscr{U}$. Let us assume that all of the subsets have size less than m. This means that $|\bigcup_{i=0}^{n-1} \mathscr{U}_i| = \sum_{i=0}^{n-1} |\mathscr{U}_i| < \sum_{i=0}^{n-1} m = nm = |\mathscr{U}|$. Therefore, $|\mathscr{U}| > |\bigcup_{i=0}^{n-1} \mathscr{U}_i|$ $\Rightarrow |\bigcup_{i=0}^{n-1} \mathscr{U}_i| \neq U$, which is a contradiction to the initial statement. Let j be such that some subset $|\mathscr{U}_j| \geq m$. If we have equality then the proof is complete, otherwise if the subset is of size greater than m, then the subset can be partitioned into two parts, one of the requisite size m, and the remainder of the subset. This gives the subset of size m, as required.
If you have to hash v values, and k keys, and for the first value of m for which $v \geq m \cdot k$, then there is a subset of size m values which hash to the same key. As such in hashing with chaining, the worst case time complexity to find a value is $\Omega(m)$.

Question 2

(a)

There are $n = 1000$ different buckets that values could be hashed into. The probability that k values are hashed to different buckets is given by the following formula $\frac{n!}{(n-k)! \cdot n^k} = \frac{k! \cdot \binom{n}{k}}{n^k}$. As such the probability that two values hash to the same key, denoted S, is $P(S) = 1 - \frac{k! \cdot \binom{n}{k}}{n^k}$. k is equal to the number of insertions. Let $k = 57$, $P(S) = 1 - \frac{57! \cdot \binom{1000}{57}}{1000^{57}} = 0.8034....$ For $k = 56$, $P(S) = 0.7918....$ Therefore, the answer is $k = 57$.

(b)

We are given that in the first $k-1$ insertions there have been no collisions and that on the k-th insertion the probability of collision must be strictly less than 20%. The first fact means that $k-1$ buckets are used in the hash table $\Rightarrow$ the probability of collision in the k-th insertion should be less than $\frac{k-1}{1000}$.

From the second fact provided, we know that $\frac{k-1}{1000} < 0.2 \Rightarrow k-1 < 200 \Rightarrow k < 201$. As such, the value of k when the size of the hash table should be increased and thus the first time that the collision probability is greater than or equal to 20%, meaning that the size of the hash table needs to be increased, is on the **200th insertion**.

Question 3

We are aiming to prove that the expected running time of an entire sequence of m operations, on a hash table of size n, is upper bounded by $m \cdot (1 + \frac{m}{2n})$. Let us define L_i as the expected length of a chain in a bucket after i operations. The worst case number of collisions for the k-th operation, for element x is where x is at the end of the chain for bucket $h(x)$, irrespective of the type of operation. This means that, $K_k = L_{k-1}$ because k-th operation does not change the expected bucket size for itself. We are told that H is a 2-universal set of has functions, as such:

$$L_k \leq \sum_{i=1}^{k} \frac{1}{n} = \frac{k}{n}$$

Using the above facts we can prove the upper bound on the expected time for m operations:

$$\sum_{i=1}^{m} E_i = \sum_{i=1}^{m} 1 + K_i = m + \sum_{i=1}^{m} K_i = m + \sum_{i=1}^{m} L_{i-1} \leq m + \sum_{i=1}^{m} \frac{i-1}{n} = m + \frac{m(m-1)}{2n} = m \cdot (1 + \frac{m-1}{2n}) \leq m \cdot (1 + \frac{m}{2n})$$

2 Information Theory

Question 4

The capacity of a channel is defined as follows: $C = \max_{p(x)} I(X;Y)$. $I(X;Y) = H(Y) - H(Y|X)$. This gives us $I(X;Y) = H(Y)$ because of the fact that the channel is deterministic; if $Y = f(X)$, then $H(Y|X) = 0$. $\therefore C = \max_{p(x)} H(Y) = \max_{p(x)} H(f(X))$. $p(x)$ is the probability distribution of, which maximises entropy when the distribution is uniform. $\therefore C = log|f(X)|$ when $p(x)$ is a uniform distribution $\square$.

Question 5

Property 2

Line 1: The sum of probabilities of all input sequences $x^n \in \mathcal{X}^n$ is equal to 1.
Line 2: The typical set $A_\varepsilon^{(n)}$ is a subset of $\mathcal{X}^n$ with particular properties. Therefore, it is obvious that the number of sequences in the typical set is not greater than the number of sequences in $\mathcal{X}^n$. Thus, the sum of the probabilities for this subset is ≤ 1.
Line 3: One of the properties of typical set is that $H(X) - \varepsilon \leq -\frac{1}{n}\log p(x^n) \leq H(X) + \varepsilon$. Will focus on the RHS of this inequality; $-\frac{1}{n}\log p(x^n) \leq H(X) + \varepsilon$. Rearranging this we get $p(x^n) \geq 2^{-n(H(X)+\varepsilon)}$. $\therefore$ we get the inequality shown.
Line 4: The value $2^{-n(H(X)+\varepsilon)}$ is constant so we can pull it out of the sum $\sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)} \sum_{x^n \in A_\varepsilon^{(n)}} 1 = 2^{-n(H(X)+\varepsilon)} |A_\varepsilon^{(n)}|$. To complete the proof simply multiply both sides of the inequality by $2^{n(H(X)+\varepsilon)}$. This leaves $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$, as required.

Property 3

Line 1: The first line makes use of Property 1 of the AEP theorem for typical sequences. It states that, for n large enough, the probability of typical set is nearly 1. Should be noted that $Pr\{A_\varepsilon^{(n)}\} = \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)$.
Line 2: One of the properties of typical set is that $H(X) - \varepsilon \leq -\frac{1}{n}\log p(x^n) \leq H(X) + \varepsilon$. Will now focus on the LHS of this inequality; $-\frac{1}{n}\log p(x^n) \geq H(X) - \varepsilon$. Rearranging this we get $p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$. $\therefore$ we get the inequality shown.
Line 3: The value $2^{-n(H(X)-\varepsilon)}$ is constant so we can pull it out of the sum $\sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)} \sum_{x^n \in A_\varepsilon^{(n)}} 1 = 2^{-n(H(X)-\varepsilon)} |A_\varepsilon^{(n)}|$. To complete the proof simply multiply both sides of the inequality by $2^{n(H(X)-\varepsilon)}$. This leaves $|A_\varepsilon^{(n)}| \geq (1-\varepsilon)2^{n(H(X)+\varepsilon)}$, as required.

Question 6

(a)

Let $Z = S \cap T, S = Z \cup S', T = Z \cup T'$. Given the definitions, S', T', and Z are disjoint.

$$\begin{aligned}
H(X_{S\cup T}) + H(X_{S\cap T}) &= H(X_Z, X_{S'}, X_{T'}) + H(X_Z) \\
&= 2H(X_Z) + H(X_{S'}, X_{T'}|X_Z) \\
&= 2H(X_Z) + H(X_{S'}|X_Z) + H(X_{T'}|X_Z) - I(X_{S'}, X_{T'}|X_Z) \\
&\leq H(X_Z) + H(X_{S'}|X_Z) + H(X_Z) + H(X_{T'}|X_Z) \\
&= H(X_Z, X_{S'}) + H(X_Z, X_{T'}) \\
&= H(X_S) + H(X_T) \ \square
\end{aligned}$$

(b)

$$\sum_{i=1}^{m} H(X_{[m]/\{i\}}) = H(X_2,...,X_m) + H(X_1,X_3,...,X_m) + ... + H(X_1,...,X_{m-1})$$

$$= \sum_{i=1}^{m} H(X_i|X_{i-1},...,X_2) - H(X_1) + \sum_{i=1}^{m} H(X_i|X_{i-1},...,X_3,X_1)$$

$$-H(X_2|X_1) + ... + \sum_{i=1}^{m} H(X_i|X_{i-1},...,X_1) - H(X_m|X_{m-1},...,X_1)$$

Conditioning reduces entropy which means that adding the respective missing discrete random variable to RHS of each conditional entropy reduces the overall entropy.

$$\therefore \sum_{i=1}^{m} H(X_{[m]/\{i\}}) \geq m\sum_{i=1}^{m} H(X_i|X_{i-1},...,X_1) - H(X_1) - H(X_2|X_1) - ... - H(X_m|X_{m-1},...,X_1)$$

$$= m\sum_{i=1}^{m} H(X_i|X_{i-1},...,X_1) - \sum_{i=1}^{m} H(X_i|X_{i-1},...,X_1)$$

$$= (m-1)\sum_{i=1}^{m} H(X_i|X_{i-1},...,X_1)$$

$$= (m-1)H(X_1,...,X_m) = (m-1)H(X_{[m]}).$$

Thus, $\sum_{i=1}^{m} H(X_{[m]/\{i\}}) \leq (m-1)H(X_{[m]})$ $\square$.