Final Exam Practice Questions - 10-605

Dec 8, 2016

10-605 **Name:**________________________
Fall 2016
Final Exam Practice QuestionsAndrew ID:________________________
Time Limit: 80 Minutes

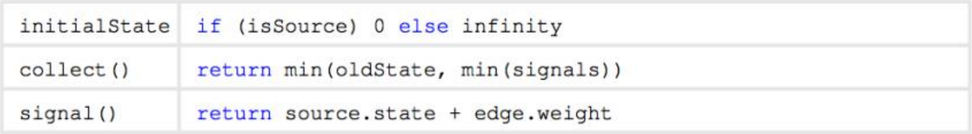
Grade Table (for teacher use only)

Question	Points	Score
1	4	
2	12	
3	12	
4	14	
5	6	
6	4	
7	8	
Total:	60	

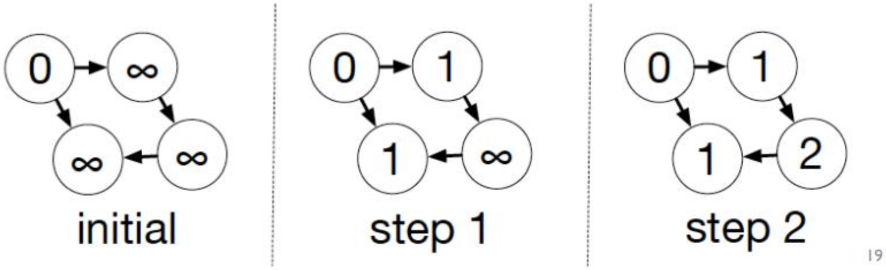
1. (4 points) True or false. Circle the correct answer.
   (a) A count-Min Sketch could be used to store feature weights for the Perceptron Algorithm. **True False**
   (b) To implement an iterative algorithm, Hadoop is generally more suitable than Spark
   True False

2. (12 points) Write a signal/collect program to detect a cycle in a graph. You will also generate a condition that, at the end of the computation, should be true for all vertices if and only if there are no loops in the graph. Note the graph given below is just an example; it does not have a cycle in the graph. You can use set operations if you wish, for example s = singleton(x) and s = union(s_1, s_2) - and assume that a unique identifier for each node is accessible with the function getId().

   For your reference we have provided you with the code for single-source shortest paths.



initialState	if (isSource) 0 else infinity
collect()	return min(oldState, min(signals))
signal()	return source.state + edge.weight



initialState ______________________________

collect ______________________________

signal ______________________________

condition to check ______________________________

3. (12 points) Randomized algorithms

   (a) (2 points) Like a HashSet, we can enumerate the elements in a Bloom Filter
   True / False

   (b) (2 points) A Bloom Filter can generate
   False Positive / False Negative answers to a query "has item x been stored in the past?" but never generate **False Positive / False Negative** answers.

   (c) (2 points) A Count-Min Sketch never *underestimates* the value associated with an element
   True / False

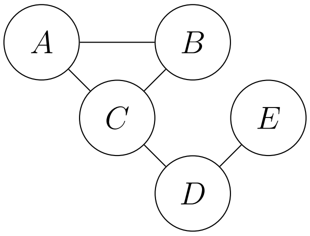
   (d) (2 points) If more bits are used for Locality Sensitive Hashing, the approximate cosine similarity will be more accurate.
   True / False

   (e) (4 points) Denote the Bloom Filter generated by inserting all the elements of S as $BF(S)$, and the value of i-th bucket of a Bloom Filter X as $X[i]$. Suppose we always use the same set of hash functions and the same number of buckets m. Circle the correct answer:

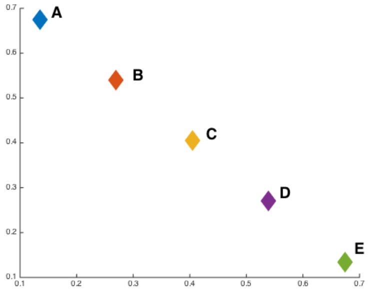
   - Let $C[i] = (BF(S)[i] \text{ OR } BF(T)[i]), \forall 1 \leq i \leq m$.
     Then $C = BF(S \cup T)$. **True / False**
   - Let $D[i] = (BF(S)[i] \text{ XOR } BF(T)[i]), \forall 1 \leq i \leq m$.
     Then $D = BF(S \cap T)$. **True / False**

4. (14 points) Spectral clustering.

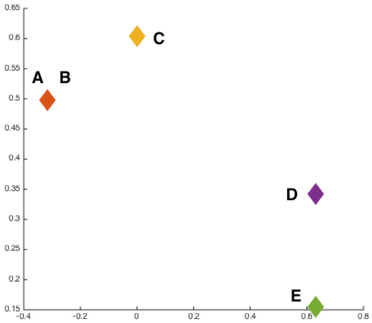
(a) (8 points) You are given the graph below and need to use spectral clustering to divide the vertices into two groups.



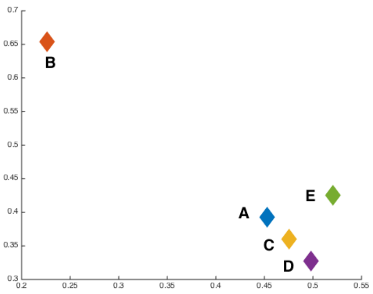
Below are four possible 2D visualization of the first two eigenvectors of the normalized adjacency matrix of the graph above. Which might be the correct one? Please circle the correct letter(s) (A, B, C, D). Optionally, add 1-2 sentences explaining your reasoning.



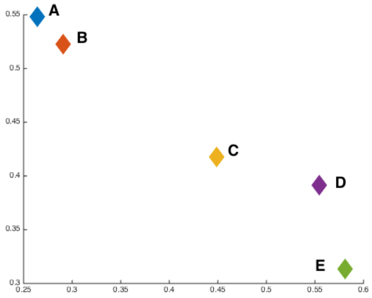
A



B



C



D

5. (6 points) **Latent Dirichlet Allocation**

Here is the list of notations used for this question:
$\boldsymbol{M}$ — number of documents
$\boldsymbol{N}_d$ — number of word tokens in document m
$\boldsymbol{K}$ — number of topics
$\boldsymbol{V}$ — number of terms in vocabulary
$\vec{\boldsymbol{\alpha}}$ — hyperparameter ($\boldsymbol{K}$-dimensional vector)
$\vec{\boldsymbol{\beta}}$ — hyperparameter ($\boldsymbol{V}$-dimensional vector)

(a) (6 points) Fill in the **two** blanks below to complete the generative process for LDA.

```
**for** all topics $t \in [1, \boldsymbol{K}]$ **do**
    sample a topic distribution over terms: $\vec{\varphi}_t \sim$ Dirichlet($\vec{\boldsymbol{\beta}}$)
**end for**
**for** all documents $d \in [1, \boldsymbol{M}]$ **do**
    sample distribution over topics in document $d$: $\vec{\theta}_d \sim$ ______________
    **for** all word positions $i \in [1, \boldsymbol{N}_d]$ in document $d$ **do**
        sample topic index $z_{d,i} \sim$ Multinomial($\vec{\theta}_d$)
        sample word $w_{d,i} \sim$ ______________
    **end for**
**end for**
```

6. (4 points) **SSL on Graphs**

   (a) (2 points) Which algorithm(s) will often downweight noisy seeds?

   (A) Harmonic Field

   (B) MultiRank Walk

   (C) Both of the first two answers

   (D) Neither of the first two answers

   (b) (2 point) Let n be the number of nodes in the data, m be the number of labels, and e be the number of edges. In the expanded version of MAD (MAD-Sketch), how many Count-Min Sketches are used to represent the graph?

   (A) $O(n)$

   (B) $O(m)$

   (C) $O(e)$

   (D) 1

7. (8 points) **Parameter Server Concepts**

   (a) (4 points) Circle the correct answer:
   In a Parameter Server, a **worker / server** node usually processes data and executes machine learning algorithms, while a **worker / server** node stores and synchronizes parameters

   A Parameter Server supports two kinds of parallelism for machine learning algorithms. **Data / Model** Parallelism partitions data to workers, while **Data / Model** Parallelism partitions parameters.

   (b) (2 points) Circle True or False:
   With a Parameter Server, we can *not* perform Data Parallelism and Model Parallelism strategies simultaneously.
   True **False**

   (c) (2 points) Circle true or false:
   Compared to a fully synchronous Parameter Server, SSP with $\tau > 0$ will generally require *fewer* epochs for the algorithm to converge.
   True **False**