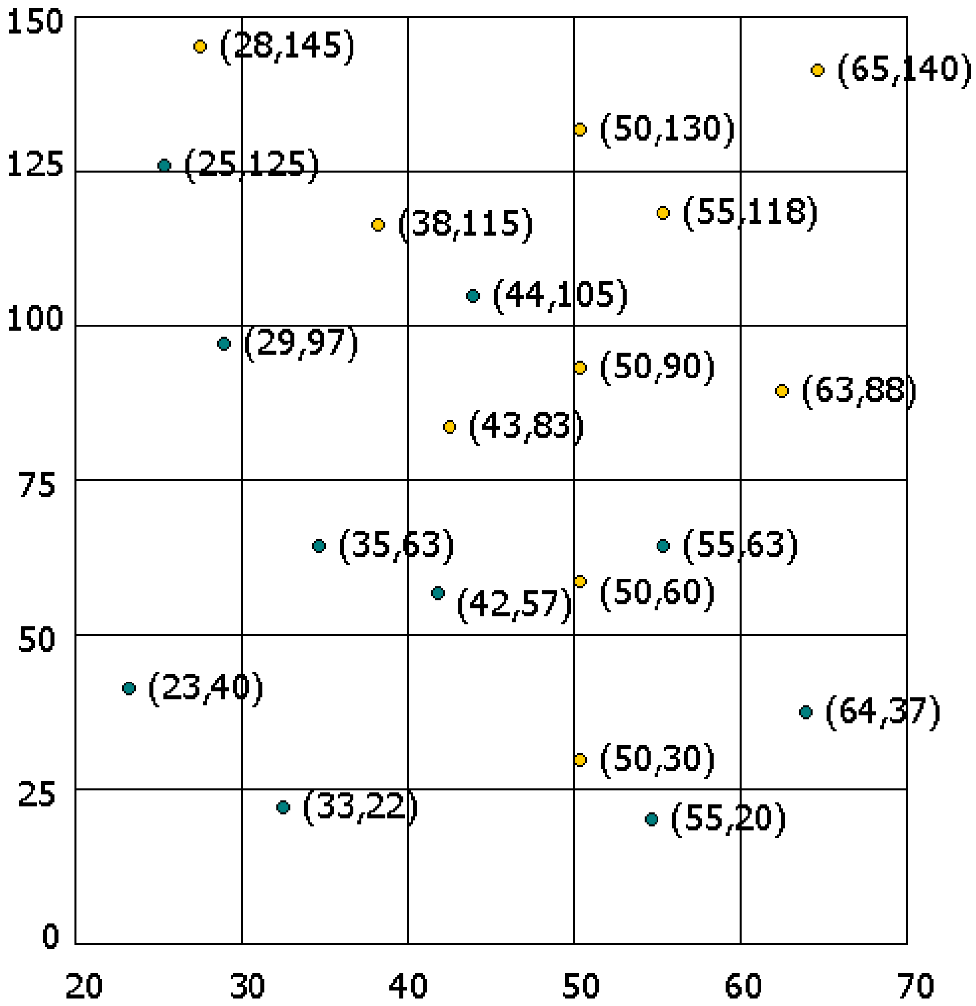
Feedback — Week5B (Basic)

Help Center

You submitted this quiz on **Wed 4 Mar 2015 6:01 PM PST**. You got a score of **5.00** out of **5.00**.

Question 1

We wish to cluster the following set of points:



into 10 clusters. We initially choose each of the green points (25,125), (44,105), (29,97), (35,63), (55,63), (42,57), (23,40), (64,37), (33,22), and (55,20) as a centroid. Assign each of the gold points to their nearest centroid. (Note: the scales of the horizontal and vertical axes differ, so you really need to apply the formula for distance of points; you can't just "eyeball" it.) Then, recompute the centroids of each of the clusters. Do any of the points then get reassigned to a new cluster on the next round? Identify the true statement in the list below. Each statement refers either to a centroid AFTER recomputation of centroids (precise to one decimal place) or to a point that gets reclassified.

Your Answer		Score	Explanation
There is a centroid after recomputation at (51.1,105.5)			
There is a centroid after recomputation at (55,20)			
There is a centroid after recomputation at (52.5,109.3)	✔	1.00	
There is a centroid after recomputation at (50.3,116.3)			
Total		1.00 / 1.00	

Question Explanation

(28,145) and (50,130) are assigned to the cluster of (25,125), so the new centroid for this cluster is (34.3,133,3).
(43,83) is assigned to the cluster of (29,97). No other points are assigned to this cluster, so the new centroid for this cluster is (36,90).

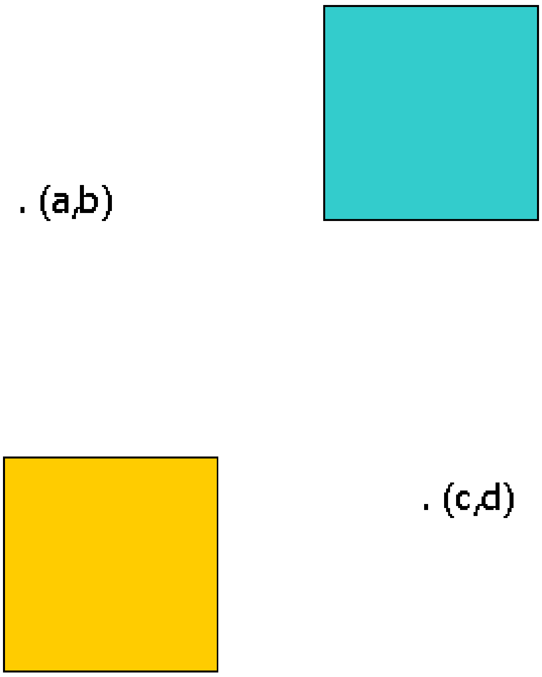
(50,60) is assigned to the cluster of (55,63). No other points are assigned to this cluster, so the new centroid for this cluster is (52.5,61.5).

(50,30) is assigned to the cluster of (55,20). No other points are assigned to this cluster, so the new centroid for this cluster is (52.5,25).

The remaining five gold points are assigned to the cluster of (44,105). The centroid of the six points in this cluster is (52.5,109.3).

Question 2

When performing a k-means clustering, success depends very much on the initially chosen points. Suppose that we choose two centroids (a,b) = (5,10) and (c,d) = (20,5), and the data truly belongs to two rectangular clusters, as suggested by the following diagram:

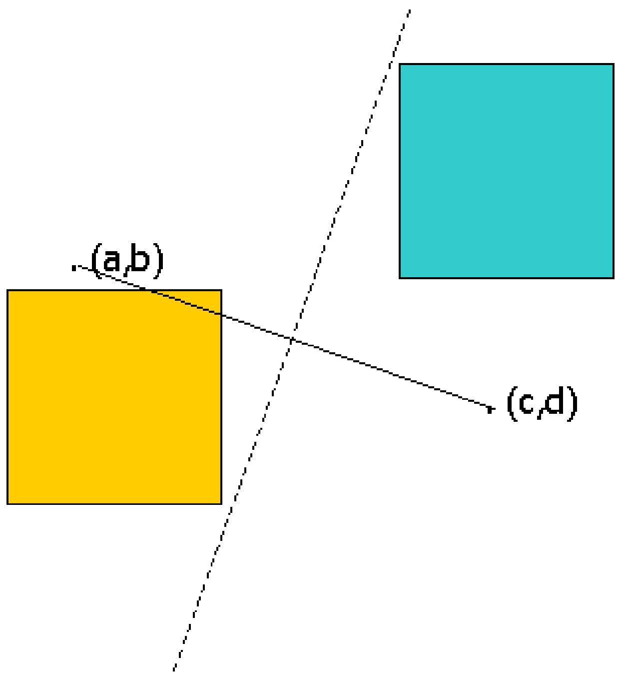


Under what circumstances will the initial clustering be successful? That is, under what conditions will all the yellow points be assigned to the centroid (5,10), while all of the blue points are assigned to cluster (20,5))? Identify in the list below, a pair of rectangles (described by their upper left corner, UL, and their lower-right corner LR) that are successfully clustered.

Your Answer		Score	Explanation
Yellow: UL=(7,8) and LR=(12,5); Blue: UL=(15,14) and LR=(20,10)			
Yellow: UL=(6,7) and LR=(11,4); Blue: UL=(11,5) and LR=(17,2)			
Yellow: UL=(3,3) and LR=(10,1); Blue: UL=(15,14) and LR=(20,10)	✔	1.00	
Yellow: UL=(3,15) and LR=(13,7); Blue: UL=(14,10) and LR=(23,6)			
Total		1.00 / 1.00	

Question Explanation

The key observation is that all points to the left of the perpendicular bisector of the line from (a,b) = (5,10) to (c,d) = (20,5) will be clustered with (a,b), while those to the right will be clustered with (c,d). Examine the diagram below:



Notice that for all the yellow points to lie to the left of the perpendicular bisector, it is both necessary and sufficient that the lower-right corner of the yellow rectangle is to the left of that line. Similarly, it is necessary and sufficient that the upper-right corner of the blue rectangle lie to the right of that line.

Suppose (x,y) is the lower-right corner of the yellow. Then the distance between (x,y) and (5,10) is sqrt($(x-5)^2+(y-10)^2$), and the distance from (x,y) to (20,5) is sqrt($(x-20)^2+(y-5)^2$). Since we want (x,y) to be closer to (5,10), we need $(x-5)^2+(y-10)^2 < (x-5)^2+(y-10)^2$, or $x < 10+y/3$. Likewise, for the upper-left corner of the blue to be to the right of the bisector, we need the opposite, $x > 10+y/3$ when (x,y) is the upper-left corner.

Question 3

Suppose we apply the BALANCE algorithm with bids of 0 or 1 only, to a situation where advertiser A bids on query words x and y, while advertiser B bids on query words x and z. Both have a budget of $2. Identify in the list below a sequence of four queries that will certainly be handled optimally by the algorithm.

Your Answer		Score	Explanation
xyyy			
yzyy	✔	1.00	Note that the optimum only yields $3.
yxxz			
xxxz			

Total	1.00 / 1.00

Question Explanation

The explanations are associated with the various incorrect choices.

Question 4

The set cover problem is: given a list of sets, find a smallest collection of these sets such that every element in any of the sets is in at least one set of the collection. As we form a collection, we say an element is covered if it is in at least one set of the collection. Note: In this problem, we shall represent sets by concatenating their elements, without brackets or commas. For example, {A,B} will be represented simply as AB. There are many greedy algorithms that could be used to pick a collection of sets that is close to as small as possible. Here are some that you will consider in this problem. Dumb: Select sets for the collection in the order in which they appear on the list. Stop when all elements are covered. Simple: Consider sets in the order in which they appear on the list. When it is considered, select a set if it has at least one element that is not already covered. Stop when all elements are covered. Largest-First: Consider sets in order of their size. If there are ties, break the tie in favor of the one that appears first on the list. When it is considered, select a set if it has at least one element that is not already covered. Stop when all elements are covered. Most-Help: Consider sets in order of the number of elements they contain that are not already covered. If there are ties, break the tie in favor of the one that appears first on the list. Stop when all elements are covered. Here is a list of sets: AB, BC, CD, DE, EF, FG, GH, AH, ADG, ADF First, determine the optimum solution, that is, the fewest sets that can be selected for a collection that covers all eight elements A,B,...,H. Then, determine the sizes of the collections that will be constructed by each of the four algorithms mentioned above. Compute the ratio of the size returned by the algorithm to the optimum size, and identify one of these ratios in the list below, correct to two decimal places.

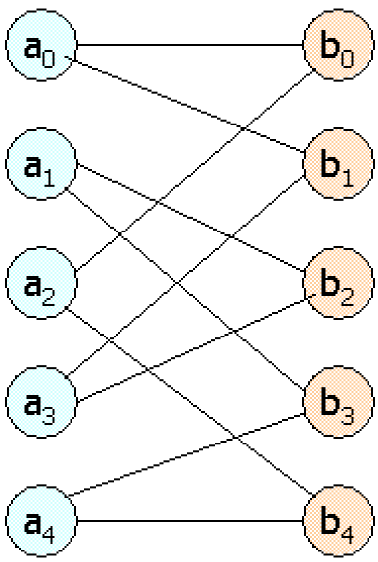
Your Answer		Score	Explanation
The ratio for Simple is 2.00			
The ratio for Most-Help is 1.00	✔	1.00	
The ratio for Most-Help is 1.25			
The ratio for Largest-First is 1.25			
Total		1.00 / 1.00	

Question Explanation

First, there are several collections of four sets that cover all elements. An example is {AB, CD, EF, GH}. There can be no smaller collection. To see why, if there are only three sets in the collection, then the sum of the sizes of these three sets must be at least 8, so two of the three sets must be ADG and ADF. But only four elements are covered by these two sets, so at least two more sets must be in the collection. Algortithm Dumb picks the first seven sets, AB, BC, CD, DE, EF, FG, GH. At that point, all eight elements are covered, so it stops with a collection of size 7. The ratio for Dumb is therefore 1.75. Algorithm Simple considers the same sets, and accepts each one, because each has an element not yet covered. Thus, its ratio is also 1.75. Algorithm Largest-First begins by taking ADG for the collection. It then considers ADF, and adds it to the collection because F is not yet covered. Then, it considers the sets of size 2 in the order listed. AB is accepted because it covers B. BC is accepted because it covers C. CD is not accepted because C and D are both covered already. DE is accepted because it covers E. Then EF and FG are not accepted, but GH is accepted because it covers H. Thus, the collection constructed by Largest-First is {ADG, ADF, AB, BC, DE, GH}, and its ratio is 1.50. Algorithm Most-Help also begins by taking ADG for the collection. But now there are two sets that add 2 to the collection: BC and EF. BC is taken first, because it precedes EF on the list, but then EF is taken. At this point, only H is missing, and the first set on the list that has H is GH, so that set is taken next. The collection constructed by Most-Help is {ADG, BC, EF, GH}, and its ratio is 1.00.

Question 5

This bipartite graph:



Has several perfect matchings. Find all the perfect matchings and then identify, in the list below, a pair of edges that can appear together in a perfect matching.

Your Answer	Score	Explanation
○ a_4-b_4 and a_3-b_1		

◉ a_3-b_2 and a_4-b_4	✔	1.00
○ a_1-b_2 and a_0-b_1		
○ a_1-b_2 and a_2-b_0		
Total		1.00 / 1.00

Question Explanation

Suppose a_0 is matched with b_0. Then a_2 can only match with b_4. So a_4 is forced to match with b_3. Next, a_1 is forced to match with b_2, and finally, a_3 is forced to match with b_1. That gives us one perfect matching.
Conversly, suppose a_0 is matched with b_1. Then a_3 can only match with b_2. So a_1 is forced to match with b_3. Next, a_4 is forced to match with b_4, and finally, a_2 is forced to match with b_0. Therefore, there are only two perfect matchings --- this one and the one in the previous paragraph.