Team 2 – Jarrett Arredondo, Krystal Grant, Colin Mettler, Chloé Plasse March 21, 2023
CS534-S23-S01 Group Project Assignment #3

**Question 1.** ***Convert the following set of sentences into the conjunctive normal form (CNF).***

S1: $A \Leftrightarrow (B \lor E)$
Biconditional elimination: $(A \Rightarrow B \lor E) \land (B \lor E \Rightarrow A)$
Implication elimination: $(\neg A \lor B \lor E) \land (\neg (B \lor E) \lor A)$
De Morgan: $(\neg A \lor B \lor E) \land (\neg B \land \neg E) \lor A)$
Distributivity Law: $(\neg A \lor B \lor E) \land (\neg B \lor A) \land (\neg E \lor A)$
CNF: $(\neg A \lor B \lor E) \land (\neg B \lor A) \land (\neg E \lor A)$

S2: $E \Rightarrow D$
Implication elimination: $\neg E \lor D$
CNF: $\neg E \lor D$

S3: $C \land F \Rightarrow \neg B$
Implication elimination: $\neg(C \land F) \lor \neg B$
De Morgan: $(\neg C \lor \neg F) \lor \neg B$
CNF: $\neg C \lor \neg F \lor \neg B$

**Question 2.** ***Assuming predicates Parent(p, q) and Female(p) and constants Joan and Kevin, with the obvious meanings, express each of the following sentences in first-order logic. (You may use the abbreviation $\exists^1$ to mean "there exists exactly one.")***

1. *Joan has a daughter (possibly more than one, and possibly sons as well).*
   ∃x Female(x) ∧ Parent(Joan,x)
   There exists an x such that x is a female and Joan is the parent of x.

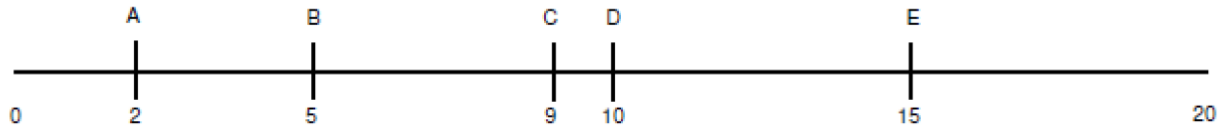
2. *Joan has exactly one daughter (but may have sons as well).*
   $\exists^1$x Female(x) ∧ Parent(Joan,x)
   There exists exactly one x such that x is a female and Joan is the parent of x.

3. *Joan has exactly one child, a daughter.*
   $\exists^1$x Parent(Joan,x) ∧ Female(x)
   There exists exactly one x such that Joan is the parent of x and x is a female.

4. *Joan and Kevin have exactly one child together.*
   $\exists^1$x Parent(Joan,x) ∧ Parent(Kevin,x)
   There exists exactly one x such that Joan is the parent of x and Kevin is the parent of x.

5. *Joan has at least one child with Kevin, and no children with anyone else.*
   (∃x Parent(Joan,x) ∧ Parent(Kevin,x)) ∧ (¬∃y Parent(Joan,y) ∧ ¬Parent(Kevin,y))
   There exists an x such that Joan is the parent of x and Kevin is the parent of x and there doesn't exist a y such that Joan is the parent of y and Kevin is not the parent of y.

**Question 3.** ***Five Customers' Rating on a New Car on a 20-point Scale.***




*(a) Assume K = 2 and the two initial centroids are 3 and 4.*

| Customer | Rating | Distance to Centroid 1 | Distance to Centroid 2 | | | Distance to Centroid 1 | Distance to Centroid 2 | | | Distance to Centroid 1 | Distance to Centroid 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 2 | 1 | 2 | | | 0 | 7.75 | | | 1.5 | 9.33 |
| B | 5 | 2 | 1 | | | 3 | 4.75 | | | 1.5 | 6.33 |
| C | 9 | 6 | 5 | | | 7 | 0.75 | | | 5.5 | 2.33 |
| D | 10 | 7 | 6 | | | 8 | 0.25 | | | 6.5 | 1.33 |
| E | 15 | 12 | 11 | | | 13 | 5.25 | | | 11.5 | 3.67 |
| | | | | | | | | | | | |
| Centroid 1 | 3 | | | New Centroid 1 | 2 | | | New Centroid 1 | 3.5 | | |
| Centroid 2 | 4 | | | New Centroid 2 | 9.75 | | | New Centroid 2 | 11.33333 | | |

1. Clusters = {2,5} and {9,10,15}
2. Silhouette Coefficient

| Customer | Rating | 2 | 5 | 9 | 10 | 15 | Intra-Cluster Distance [a(i)] | Inter-Cluster Distance [b(i)] | Silhouette Coefficient [s(i)] |
|---|---|---|---|---|---|---|---|---|---|
| A | 2 | | 3 | 7 | 8 | 13 | 3 | 9.33 | 0.68 |
| B | 5 | -3 | | 4 | 5 | 10 | 3 | 6.33 | 0.53 |
| C | 9 | -7 | -4 | | 1 | 6 | 3.5 | 5.5 | 0.36 |
| D | 10 | -8 | -5 | -1 | | 5 | 3 | 6.5 | 0.54 |
| E | 15 | -13 | -10 | -6 | -5 | | 5.5 | 11.5 | 0.52 |
| | | | | | | | | | 0.53 |

Davies-Bouldin Index

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | ci | cj | dij | | | | |
| **Customer** | **Rating** | **3.5** | **11.33** | 7.83 | | | | |
| **A** | 2 | 1.5 | | | | | | |
| **B** | 5 | 1.5 | | | | | | |
| **C** | 9 | | 2.33 | | | | | |
| **D** | 10 | | 1.33 | | | | | |
| **E** | 15 | | 3.67 | | | | | |
| | | si | sj | | Rij | Di | Dj | DB |
| | | *1.5* | *2.443333* | | 0.503619 | 0.503619 | 0.503619 | 0.503619 |

Calinski-Harabasz Index

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ci | cj | c | | | | | |
| **Customer** | **Rating** | **3.5** | **11.33** | **8.2** | | | | | |
| **A** | 2 | 1.5 | | | 44.18 | k=1 | 2.25 | i=1 | |
| **B** | 5 | 1.5 | | | | | 2.25 | i=2 | |
| **C** | 9 | | 2.33 | | 29.3907 | k=2 | 5.4289 | i=3 | |
| **D** | 10 | | 1.33 | | | | 1.7689 | i=4 | |
| **E** | 15 | | 3.67 | | | | 13.4689 | i=5 | |
| | | | | | Sum of nk \|ck - c\|^2 / K -1 for each cluster | | Sum of \|di - ck\|^2 for each point for each cluster / N - K | | CH |
| | | | | | 73.5707 | | 8.3889 | | 8.770006 |

*(b) Assume K = 2 and the two initial centroids are 11 and 12.*

| Customer | Rating | Distance to Centroid 1 | Distance to Centroid 2 | | | Distance to Centroid 1 | Distance to Centroid 2 |
|---|---|---|---|---|---|---|---|
| **A** | 2 | 9 | 10 | | | 4.5 | 13 |
| **B** | 5 | 6 | 7 | | | 1.5 | 10 |
| **C** | 9 | 2 | 3 | | | 2.5 | 6 |
| **D** | 10 | 1 | 2 | | | 3.5 | 5 |
| **E** | 15 | 4 | 3 | | | 8.5 | 0 |
| | | | | | | | |
| **Centroid 1** | 11 | | | **New Centroid 1** | 6.5 | | |
| **Centroid 2** | 12 | | | **New Centroid 2** | 15 | | |

1. Clusters = {2,5,9,10} and {15}
2. Silhouette Coefficient

| **Customer** | **Rating** | 2 | 5 | 9 | 10 | 15 | Intra-Cluster Distance [a(i)] | Inter-Cluster Distance [b(i)] | Silhouette Coefficient [s(i)] | |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | 2 | | 3 | 7 | 8 | 13 | 6 | 13.00 | 0.54 | |
| **B** | 5 | -3 | | 4 | 5 | 10 | 4 | 10.00 | 0.60 | |
| **C** | 9 | -7 | -4 | | 1 | 6 | 4 | 6.00 | 0.33 | |
| **D** | 10 | -8 | -5 | -1 | | 5 | 4.67 | 5.00 | 0.07 | |
| **E** | 15 | -13 | -10 | -6 | -5 | | 0 | 8.5 | 0.00 | since \|Ci\| = 1 |
| | | | | | | | | | **0.31** | |

Davies-Bouldin Index

| | | ci | cj | dij | | | | |
|---|---|---|---|---|---|---|---|---|
| **Customer** | **Rating** | **6.5** | **15** | 8.5 | | | | |
| **A** | 2 | 4.5 | | | | | | |
| **B** | 5 | 1.5 | | | | | | |
| **C** | 9 | 2.5 | | | | | | |
| **D** | 10 | 3.5 | | | | | | |
| **E** | 15 | | 0 | | | | | |
| | | si | sj | | Rij | Di | Dj | DB |
| | | 3 | 0 | | 0.352941 | 0.352941 | 0.352941 | 0.352941 |

Calinski-Harabasz Index

| | | ci | cj | c | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Customer** | **Rating** | **6.5** | **15** | **8.2** | | | | | |
| **A** | 2 | 4.5 | | | 11.56 | | 20.25 | | |
| **B** | 5 | 1.5 | | | | | 2.25 | | |
| **C** | 9 | 2.5 | | | | | 6.25 | | |
| **D** | 10 | 3.5 | | | | | 12.25 | | |
| **E** | 15 | | 0 | | 46.24 | | 0 | | |
| | | | | | Sum of nk \|ck - c\|^2 / K -1 for each cluster | | Sum of \|di - ck\|^2 for each point for each cluster / N - K | | CH |
| | | | | | 57.8 | | 13.66666667 | | 4.229268 |

(c) *Use the results from (a) and (b) to determine which two-cluster solution should be chosen. Please describe and explain your answer in detail.*

Comparing the Silhouette Coefficient Index, Davies–Bouldin Index, and Calinski-Harabasz Index of each set of clusters, it would appear that {2,5} and {9,10,15} should be chosen. This cluster solution offers both a higher Silhouette and Calinski-Harabasz Index. The Davies-Bouldin Index is skewed toward for {2,5,9,10} and {15} since the average distance to the centroid for the {15} cluster is 0 lowering the DB Index below that of {2,5} and {9,10,15}. The {2,5} and {9,10,15} option also offers a lower SSE value of 25.16 compared to 41 for {2,5,9,10} and {15}.