**Problem 1: Inference**
## Part 1
Facts:

x

q

Rules:

$(x \wedge y) \Rightarrow a$

$x \Rightarrow w$

$(x \wedge z) \Rightarrow b$

$(y \wedge w) \Rightarrow c$

$(q \wedge z \wedge a) \Rightarrow d$

$q \Rightarrow y$

$(a \wedge b \wedge c \wedge d) \Rightarrow e$

$w \Rightarrow z$


## Part 2
**Iteration 1:**

q, q => y, y.

x, x => w, w.


**Iteration 2:**

x, y, $(x \wedge y) \Rightarrow a$, a.

w, w => z, z.

y, w, $(y \wedge w) \Rightarrow c$.


**Iteration 3:**

x, z, $(x \wedge z) \Rightarrow b$.

q, z, a, $(q \wedge z \wedge a) \Rightarrow d$, d.


**Iteration 4:**

a, b, c, d, $(a \wedge b \wedge c \wedge d) \Rightarrow e$


## Part 3
1. [e], $((a \wedge b \wedge c \wedge d) \Rightarrow e)$ => [a, b, c, d]
2. [a, b, c, d], $((x \wedge y) \Rightarrow a)$ => [x, y, b, c, d]
3. [x, y, b, c, d], x => [y, b, c, d]
4. [y, b, c, d], $(q \Rightarrow y)$ => [q, b, c, d]

5. [q, b, c, d], q  =>  [b, c, d]
6. [b, c, d], ((x ^ z) => b)  =>  [z, c, d]
7. [z, c, d], (w => z)  =>  [w, c, d]
8. [w, c, d], (x => w)  =>  [c, d]
9. [c, d], ((y ^ w) => c)  =>  [d]
10. [d], ((q ^ z ^ a) => d)  =>  []
11. []

## Part 4
Facts:

x

q

Rules:

!x v !y v a

!x v w

!x v !z v b

!y v !w v c

!q v !z v !a v d

!q v y

!a v !b v !c v !d v e

!w v z


## Part 5
1. (!x v w) ^ x => w
2. (!w v z) ^ w => z
3. (!x v !y v a) ^ x => !y v a
4. (!q v y) ^ q => y
5. (!y v a) ^ y => a
6. (!x v !z v b) ^ x => !z v b
7. (!z v b) ^ z => b
8. (!q v !z v !a v d) ^ q => !z v !a v d
9. (!z v !a v d) ^ z => !a v d
10. (!a v d) ^ a => d
11. (!y v !w v c) ^ y => !w v c
12. (!w v c) ^ w => c
13. (!a v !b v !c v !d v !e) ^ a => !b v !c v !d v e
14. (!b v !c v !d v e) ^ b => !c v !d v e
15. (!c v !d v e) ^ c => !d v e

**16.** (!d v e) ^ d => e
**17.** e ^ !e


**Problem 2: Using sklearn to do clustering**

- **V-Measure:** A clustering evaluation metric that combines homogeneity and completeness into a single score by taking their harmonic mean.
- **Adjusted Rand Index (ARI):** A statistic used to compare the similarity between two clusterings, adjusting for chance grouping. It quantifies the agreement between predicted and true clusters, with scores ranging from –1 (complete disagreement) to 1 (perfect agreement).
- **Silhouette Score:** A metric that measures how well each data point fits within its assigned cluster compared to other clusters. For each point, the average distance to other points in the same cluster (cohesion) is compared to the average distance to points in the nearest neighboring cluster (separation). The score for each point (and the overall average) ranges from -1 (misclustered) to 1 (well-clustered), with values near 0 indicating overlapping clusters.

| | Baseline ('alt.atheism', 'talk.religion.misc', 'comp.graphics', 'sci.space') | Four Similar Groups ('comp.graphics', 'comp.os.ms-windows.misc', 'comp.sys.ibm.pc.hardware', 'comp.sys.mac.hardware') | Four Different Groups ('rec.autos', 'alt.atheism', 'sci.space', 'talk.politics.mideast') |
|---|---|---|---|
| Vanilla | Homogeneity: 0.003 ± 0.000 Completeness: 0.170 ± 0.000 V-measure: 0.006 ± 0.000 Adjusted Rand-Index: -0.000 ± 0.000 Silhouette Coefficient: 0.931 ± 0.003 | Homogeneity: 0.003 ± 0.000 Completeness: 0.172 ± 0.001 V-measure: 0.005 ± 0.000 Adjusted Rand-Index: -0.000 ± 0.000 Silhouette Coefficient: 0.994 ± 0.001 | Homogeneity: 0.007 ± 0.004 Completeness: 0.138 ± 0.037 V-measure: 0.012 ± 0.008 Adjusted Rand-Index: 0.000 ± 0.000 Silhouette Coefficient: 0.841 ± 0.049 |

| TFIDF | Homogeneity: 0.349 ± 0.010 | Homogeneity: 0.120 ± 0.010 | Homogeneity: 0.313 ± 0.044 |
|---|---|---|---|
| | Completeness: 0.398 ± 0.009 | Completeness: 0.160 ± 0.012 | Completeness: 0.410 ± 0.034 |
| | V-measure: 0.372 ± 0.009 | V-measure: 0.137 ± 0.011 | V-measure: 0.355 ± 0.041 |
| | Adjusted Rand-Index: 0.203 ± 0.017 | Adjusted Rand-Index: 0.062 ± 0.008 | Adjusted Rand-Index: 0.201 ± 0.030 |
| | Silhouette Coefficient: 0.007 ± 0.000 | Silhouette Coefficient: 0.011 ± 0.001 | Silhouette Coefficient: 0.010 ± 0.001 |
| LSA | Homogeneity: 0.315 ± 0.006 | Homogeneity: 0.115 ± 0.003 | Homogeneity: 0.263 ± 0.062 |
| | Completeness: 0.400 ± 0.007 | Completeness: 0.158 ± 0.004 | Completeness: 0.416 ± 0.037 |
| | V-measure: 0.353 ± 0.007 | V-measure: 0.133 ± 0.004 | V-measure: 0.318 ± 0.058 |
| | Adjusted Rand-Index: 0.173 ± 0.013 | Adjusted Rand-Index: 0.058 ± 0.003 | Adjusted Rand-Index: 0.136 ± 0.054 |
| | Silhouette Coefficient: 0.047 ± 0.008 | Silhouette Coefficient: 0.074 ± 0.002 | Silhouette Coefficient: 0.064 ± 0.033 |

Overall, the results support our hypothesis because TFIDF and LSA show better clustering performance on the baseline and four different groups (which have more distinct classes) than the four similar groups.

**Problem 4: Knowledge Graphs**
1. **LinkedPeople uses WikiData to construct family trees for real and fictional people. Let's look at Abraham Lincoln.**
    a. **How many children did Lincoln have?**
        i. 4 children
    b. **What were their names?**
        i. Robert Todd Lincoln
        ii. Edward Baker Lincoln
        iii. William Wallace Lincoln
        iv. Tad Lincoln
    c. **Who was the father of the wife of Lincoln's son Robert?**
        i. James Harlan

2. **Entitree provides a tree-based representation for Wikidata entries, and allows you to specify what relationship you want to consider. Let's start with Lincoln again.**
   a. **What event was he the target of?**
      i. Assasination of Abraha Lincoln
   b. **Who was the perpetrator?**
      i. John Wilkes Booth
   c. **Where is the perpetrator buried?**
      i. Green Mount Cemetery
   d. **Who are three other people buried there?**
      i. Sidney Lanier
      ii. Benjamin Huger
      iii. William Pinkney Whyte
3. **OpenArtBrowser uses WikiData to provide connections and context between artworks, artists, and periods. Let's start with Leonardo Da Vinci.**
   a. **What movements is he a part of?**
      i. High Renaissance
      ii. Italian Renaissance
      iii. Renaissance
      iv. Early Renaissance
   b. **Find his work "Vitruvian Man."**
      i. **Where is it located?**
         1. Gallerie Dell'Accademia
      ii. **Who are two other artists whose works are displayed there"**
         1. Giovvani Bellini
         2. Jacopo Tintoretto
   c. **One of the motifs used in "Vitruvian Man" is the circle. What are two other artworks that use the circle as a motif?**
      i. Rhythms
      ii. Tabatinguera