Set Theory

- Sets: Definition, Universal Set, Complement, Cardinality
- Subset and Power Set
- Sets Operations
- Set Equality
- Characteristic Vectors: Sets as Bit-Vectors
- Multisets

Imdad ullah Khan

Multiset

A multiset is an unordered collection of objects where repetition of the elements matters

- A : $\{1, 2, 2, 2, 3, 3\}$
- B : $\{CS100, CS100, CS100, CS210, CS210\}$
- C : the multiset of last names of the all professors in LUMS

Order of elements is not significant

- $\{1, 2, 2, 3\}$ is the same as $\{2, 3, 1, 2\}$

Repetition counts!

- $\{1, 2, 2, 2, 3\}$ is not the same as $\{1, 2, 3\}$

Multisets: Terminology

- Multiset is also termed as bag or mset
- Number of instances of each element in a multiset is called multiplicity

An infinite number of multisets exist which contain only elements a and b, but vary in the multiplicities of their elements

- $\{a, b\}$
- $\{a, a, b\}$
- $\{a, a, a, b, b, b\}$

All of these are different multisets

All of these represent the same set

Multisets: Multiplicity

Multisets can be represented as a set of ordered pairs $\big(x, m_A(x)\big)$

- x is an element in the multiset A
- $m_A(x)$ is the multiplicity of x in the multiset A

$$\{a, a, b, b, b\} \quad \longrightarrow \quad \big\{(a, 2), (b, 3)\big\}$$

$$\{1, 2, 3, 2, 1\} \quad \longrightarrow \quad \big\{(1, 2), (2, 2), (3, 1)\big\}$$

$$\{Khan, Ali, Khan, Ali, Ayesha\} \quad \longrightarrow \quad \big\{(Ayesha, 1), (Khan, 2), (Ali, 2)\big\}$$

Multisets: Support and Cardinality

The support of a multiset A in a universe U is the underlying set of A

$$support(A) = \{x \in U \mid m_A(x) > 0\}$$

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\} \implies support(A) = \{Ayesha, Khan, Ali\}$$

The cardinality of a multiset A is the sum of multiplicities of its elements

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\}$$

$$|A| = 2 + 2 + 1 = 5$$

$$|Support(A)| = 3$$

Multisets: Inclusion

A multiset A is included in the multiset B if $\forall x \in U,\ m_A(x) \leq m_B(x)$

denoted as $A \subseteq B$

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\}$$

$$B = \{(Ayesha, 3), (Khan, 2), (Ali, 2), (Imdad, 1)\}$$

Is $A \subseteq B$?

Is $B \subseteq A$?

Multisets: Union

The union of a multiset A and a multiset B is a multiset C such that

$$\forall x \in U,\ m_C(x) = max\big(m_A(x), m_B(x)\big)$$

denoted as $C = A \cup B$

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\}$$

$$B = \{(Ayesha, 3), (Ali, 2), (Imdad, 1)\}$$

$$A \cup B = \{(Ayesha, 3), (Ali, 2), (Imdad, 1), (Khan, 2)\}$$

Multisets: Intersection

The intersection of a multiset A and a multiset B is a multiset C such that

$$\forall x \in U,\ m_C(x) = min\big(m_A(x), m_B(x)\big)$$

denoted as $C = A \cap B$

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\}$$

$$B = \{(Ayesha, 3), (Ali, 2), (Imdad, 1)\}$$

$$A \cap B = \{(Ayesha, 2), (Ali, 1)\}$$

Multisets: Sum

The sum of a multiset A and a multiset B is a multiset C such that

$$\forall x \in U,\ m_C(x) = m_A(x) + m_B(x)$$

denoted as $C = A \sqcup B$

$$A = \{(Ayesha, 2), (Khan, 2), (Ali, 1)\}$$

$$B = \{(Ayesha, 3), (Ali, 2), (Imdad, 1)\}$$

$$A \sqcup B = \{(Ayesha, 5), (Ali, 3), (Imdad, 1), (Khan, 2)\}$$

Sum of two multisets ($\sqcup$) is also known as disjoint union

Multisets: Difference

The difference of a multiset B from a multiset A is a multiset C such that

$$\forall x \in U,\ m_C(x) = max\big(m_A(x) - m_B(x), 0\big)$$

denoted as $C = A \setminus B$

$$A = \{(Ayesha, 5), (Khan, 2), (Ali, 1)\}$$
$$B = \{(Ayesha, 3), (Ali, 2), (Imdad, 1)\}$$

$$A \setminus B = \{(Ayesha, 2), (Khan, 2)\}$$

The Set-of-Words Vector Model for Text Representation

Set-of-Words: Documents represented by vectors $\in \{0,1\}^{|\Sigma|}$

	Anthony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth	...
ANTHONY	1	1	0	0	0	1	
BRUTUS	1	1	0	1	0	0	
CAESAR	1	1	0	1	1	1	
CALPURNIA	0	1	0	0	0	0	
CLEOPATRA	1	0	0	0	0	0	
MERCY	1	0	1	1	1	1	
WORSER	1	0	1	1	1	0	
...							

The Bag of words Vector Model for Text Representation

Bag-of-Words: Documents represented by term-frequency vectors $\in \mathbb{N}^{|\Sigma|}$

	Anthony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth	...
ANTHONY	157	73	0	0	0	1	
BRUTUS	4	157	0	2	0	0	
CAESAR	232	227	0	2	1	0	
CALPURNIA	0	10	0	0	0	0	
CLEOPATRA	57	0	0	0	0	0	
MERCY	2	0	3	8	5	8	
WORSER	2	0	1	1	1	5	
...							