CS 4375 Assignment 3

Jaemin Lee

Due 11/10/2019

**Section A**

**Question 1**

**Question 2**

1. A-B-C-E-D-F
2. Document 1

Vocab size: 8

Word: “machin”

Tf = 1/8 = 0.125

Idf = |3|/1+1= 3/2 = 1.5

Tf-idf = tf \* idf = 0.1875

Word: “learn”

Tf = 1/8 = 0.125

Idf = |3|/1+2 = 3/3 = 1

]tf-idf = tf \* idf = 0.125

Word: “key”

Tf = 1/8 = 0.125

Idf = |3|/1+1= 3/2 = 1.5

Tf-idf = tf \* idf = 0.1875

Word: “area”

Tf = 1/8 = 0.125

Idf = |3|/1+1 = 1.5

Tf-Idf = tf \* idf = 0.1875

Word: “focu”

Tf = 1/8 = 0/125

Idf = |3|/1+1 = 1.5

Tf-Idf = tf \* idf = 0.1875

Word: “comput”

Tf = 1/8 = 0.125

Idf = |3|/1+2 = 1

Tf-Idf = tf \* idf = 0.125

Word: “scienc”

Tf = 1/8 = 0.125

Idf = |3|/1+1 = 1.5

Tf-Idf = tf \* idf = 0.1875

Word: “research”

Tf = 1/8 = 0.125

Idf = |3|/1+2 = 1

Tf-Idf = tf \* idf = 0.125

Biggest: tie between “machin”, “key”, “area”, “focu” and “scienc”

Smallest: tie between “learn”, “comput”, and “research”

Document 2

Vocab size: 9

Word: “deep”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “learn”

Tf = 2/10 = 0.2

Idf = |3|/1+2 = 1

Tf-idf = tf\*idf = 0.2

Word: “advanc”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “method”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “involv”

Tf = 1/10 = 0.1

Idf = |3|/1+2 = 1

Tf-idf = tf \* idf = 0.1

Word: “neural”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “network”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “base”

Tf = 1/10 = 0.1

Idf = |3|/1+1 = 3/2 = 1.5

Tf-idf = tf \* idf = 0.15

Word: “comput”

Tf = 1/10 = 0.1

Idf = |3|/1+2 = 1

Tf-idf = tf \* idf = 0.1

Biggest score: “learn”, (second highest is tie between) “deep”, “advanc”, “method”, “network”, “neural”, and “base”

Lowest score: tie between “involv”, “comput”

Document 3

Vocab size: 6

Word: “ani”

Tf = 1/6 = 0.167

Idf = |3|/1+1 = 1.5

Tf-idf = tf \* idf = 0.25

Word: “involv”

Tf = 1/6 = 0.167

Idf = |3|/1+2 = 1

Tf-idf = tf \* idf = 0.167

Word: “scientif”

Tf = 1/6 = 0.167

Idf = |3|/1+1 = 1.5

Tf-idf = tf \* idf = 0.25

Word: “research”

Tf = 1/6 = 0.167

Idf = |3|/1+2 = 1

Tf-idf = tf \* idf = 0.167

Word: “help”

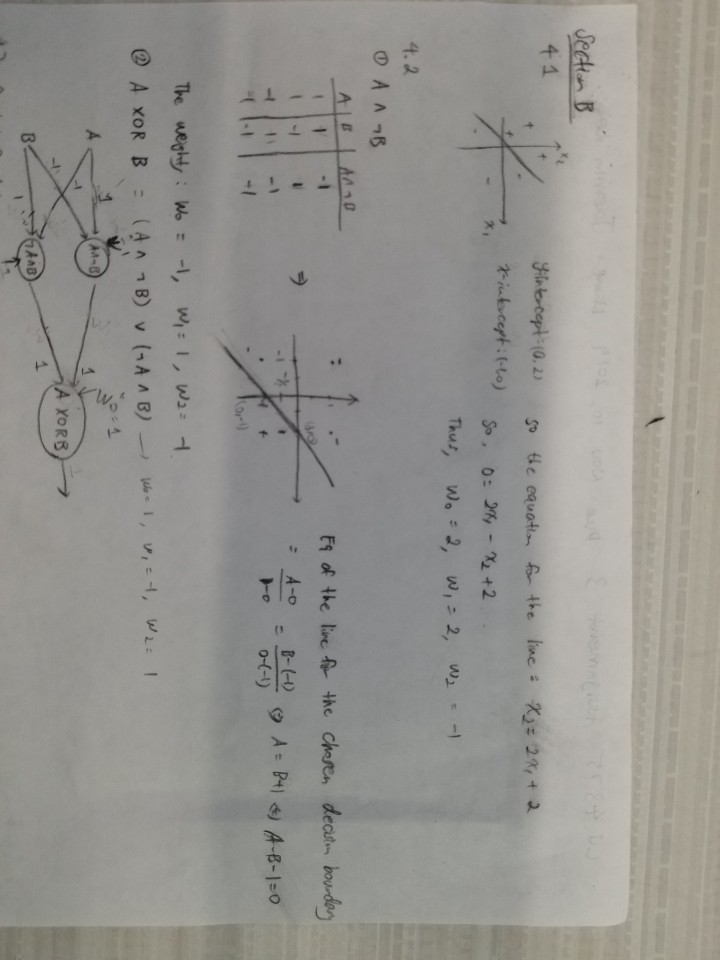
Tf = 1/6 = 0.167

Idf = |3|/1+1 = 1.5

Tf-idf = tf \* idf = 0.25

Word: “progress”

Tf = 1/6 = 0.167

Idf = |3|/1+1 = 1.5

Tf-idf = tf \* idf = 0.25

Biggest score: tie between “ani”, “scientif”, “help” and “progress”

Smallest score: tie between “involv” and “research”

**Section B**

**4.7**

