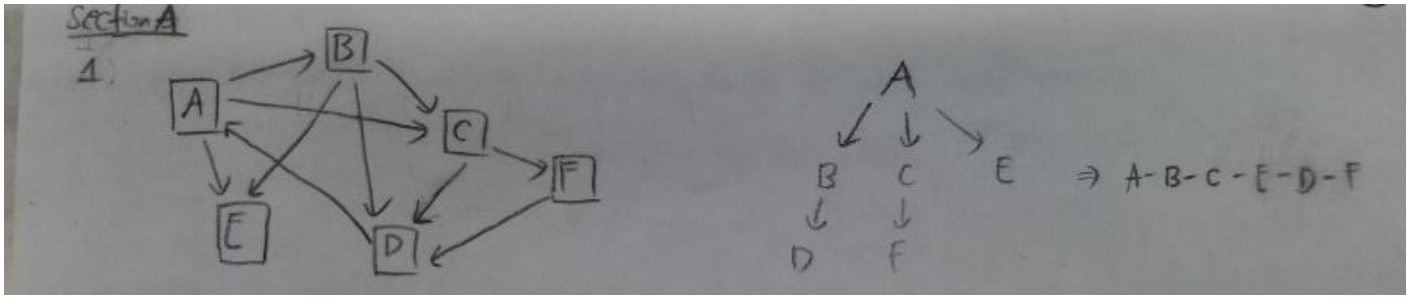
CS 4375 Assignment 3

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

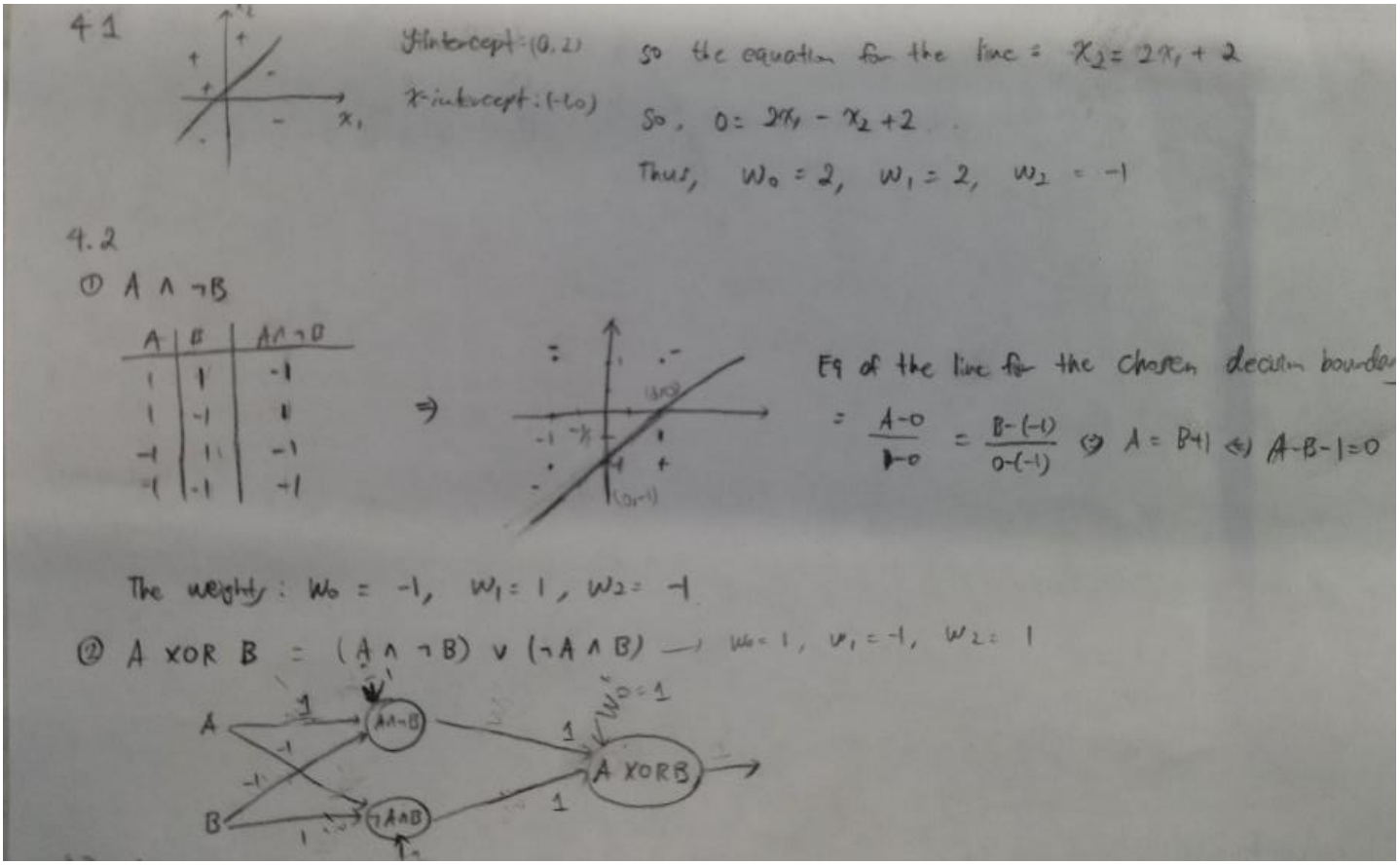
Word: “research”
Tf = 1/6 = 0.167
Idf = |3|/1+2 = 1
Tf-idf = tf * idf = 0.167

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

4.7 a = 1, b = 0, d = 1

① after first training iteration ($\eta = 0.3$, $\alpha = 0.9$, $W_{ca} = W_{cb} = W_{co} = W_{dc} = W_{do} = 0.1$):

$O_c = \sigma(W_{co} + W_{ca} \times a + W_{cb} \times b) = \sigma(0.1 + 0.1 \times 1 + 0.1 \times 0) = \sigma(0.2) = \frac{1}{1+e^{-0.2}} = 0.5498$

$O_d = \sigma(W_{do} + W_{dc} \times c) = \sigma(0.1 + 0.1 \times 0.54983) = \sigma(0.154983) = \frac{1}{1+e^{-0.154983}} = 0.53861$

$\delta_d = O_d(1-O_d)(d-O_d) = 0.53861(1-0.53860)(1-0.53861) = 0.11464$

$\delta_c = O_c(1-O_c)(W_{dc} \times \delta_d) = 0.54983(1-0.54983)(0.1 \times 0.11464) = 0.00283$

$\Delta W_{ca}(1) = \eta \cdot \delta_c \cdot a + \alpha \cdot \Delta W_{ca}(1-1) = 0.3 \cdot 0.00283 \cdot 1 + 0.9 \cdot 0 = 0.00085$

$\Delta W_{cb}(1) = \eta \cdot \delta_c \cdot b + \alpha \cdot \Delta W_{cb}(1-1) = 0.3 \cdot 0.00283 \cdot 0 + 0.9 \cdot 0 = 0$

$\Delta W_{co}(1) = \eta \cdot \delta_c \cdot x_o + \alpha \cdot \Delta W_{co}(1-1) = 0.3 \cdot 0.00283 \cdot 1 + 0.9 \cdot 0 = 0.00085$

$\Delta W_{dc}(1) = \eta \cdot \delta_d \cdot C + \alpha \cdot \Delta W_{dc}(1-1) = 0.3 \cdot 0.11464 \cdot 0.54983 + 0.9 \cdot 0 = 0.01891$

$\Delta W_{do}(1) = \eta \cdot \delta_d \cdot x_o + \alpha \cdot \Delta W_{do}(1-1) = 0.3 \cdot 0.11464 \cdot 1 + 0.9 \cdot 0 = 0.03439$

update each network weight

$W_{ca} = W_{ca} + \Delta W_{ca} = 0.1 + 0.00085 = 0.10085$

$W_{cb} = W_{cb} + \Delta W_{cb} = 0.1 + 0 = 0.1$

$W_{co} = W_{co} + \Delta W_{co} = 0.1 + 0.00085 = 0.10085$

$W_{dc} = W_{dc} + \Delta W_{dc} = 0.1 + 0.01891 = 0.11891$

$W_{do} = W_{do} + \Delta W_{do} = 0.1 + 0.03439 = 0.13439$

4.7 continued

② After the training iteration of the Back propagation algorithm:

a = 0, b = 1, d = 0

$O_c = \sigma(W_{co} + W_{ca} \times a + W_{cb} \times b) = \sigma(0.10085 + 0.10085 \times 0 + 0.1 \times 1) = \sigma(0.20085) = \frac{1}{1+e^{-0.20085}} = 0.55004$

$O_d = \sigma(W_{do} + W_{dc} \cdot C) = \sigma(0.13439 + 0.11891 \times 0.55004) = \sigma(0.19979) = \frac{1}{1+e^{-0.19979}} = 0.54978$

$\delta_d = O_d \cdot (1-O_d) \cdot (d-O_d) = 0.54978 \cdot (1-0.54978) \cdot (1-0.54978) = -0.13608$

$\delta_c = O_c \cdot (1-O_c) \cdot (W_{dc} \cdot \delta_d) = 0.55004 \cdot (1-0.55004) \cdot (0.11891 \cdot (-0.13608)) = -0.004$

$\Delta W_{ca}(2) = \eta \cdot \delta_c \cdot a + \alpha \cdot \Delta W_{ca}(2-1) = 0.3(-0.004) \cdot 0.9(0.00085) \cdot 1 = 0.00076$

$\Delta W_{cb}(2) = \eta \cdot \delta_c \cdot b + \alpha \cdot \Delta W_{cb}(2-1) = 0.3(-0.004) \cdot 1 + 0.9(0)(1) = -0.0012$

$\Delta W_{co}(2) = \eta \cdot \delta_c \cdot x_o + \alpha \cdot \Delta W_{co}(2-1) = 0.3(-0.004) \cdot 1 + 0.9(0.00085) \cdot 1 = -0.00043$

$\Delta W_{dc}(2) = \eta \cdot \delta_d \cdot C + \alpha \cdot \Delta W_{dc}(2-1) = 0.3(-0.13608) \cdot 0.55004 + 0.9(0.01891) = -0.00543$

$\Delta W_{do}(2) = \eta \cdot \delta_d \cdot x_o + \alpha \cdot \Delta W_{do}(2-1) = 0.3(-0.13608) \cdot 0 + 0.9(0.03439) \cdot 1 = -0.00987$

update each network weight

$W_{ca} = W_{ca} + \Delta W_{ca} = 0.10085 + 0.00076 = 0.10161$

$W_{cb} = W_{cb} + \Delta W_{cb} = 0.1 + (-0.0012) = 0.0988$

$W_{co} = W_{co} + \Delta W_{co} = 0.10085 + (-0.00043) = 0.10042$

$W_{dc} = W_{dc} + \Delta W_{dc} = 0.11891 + (-0.00543) = 0.11348$

$W_{do} = W_{do} + \Delta W_{do} = 0.13439 + (-0.00987) = 0.12452$