Question 1. (15 points)

The examples map from [x1, x2] to [x1, x1 x2] coordinates as follows: [−1, −1] (negative) maps to [−1, +1] [−1, +1] (positive) maps to [−1, −1] [+1, −1] (positive) maps to [+1, −1] [+1, +1] (negative) maps to [+1, +1] Thus, the positive examples have x1x2 = − 1 and the negative examples have x1x2 = + 1.

The maximum margin separator is the line x1x2 = 0, with a margin of 1. The separator corresponds to the x1 = 0 and x2 = 0 axes in the original space—this can be thought of as the limit of a hyperbolic separator with two branches.

Question 2. (5 points)

Yes. As was discussed in the lectures, diversity among the constituent models is desirable when building an ensemble predictor. This can lead to some of the individual models not exhibiting good performance considering all of the training data, but their good performance on critical subsets of the training data may result in good performance for the ensemble as a whole.

Question 3. (20 points)

(a) Answer:

$$\begin{aligned} s(\{e\}) &= \frac{8}{10} = 0.8 \\ s(\{b,d\}) &= \frac{2}{10} = 0.2 \\ s(\{b,d,e\}) &= \frac{2}{10} = 0.2 \end{aligned}$$

(b) Answer:

$$\begin{aligned} c(bd \longrightarrow e) &= \frac{0.2}{0.2} = 100\% \\ c(e \longrightarrow bd) &= \frac{0.2}{0.8} = 25\% \end{aligned}$$

(c) Answer:

$$\begin{aligned} s(\{e\}) &= \frac{4}{5} = 0.8 \\ s(\{b,d\}) &= \frac{5}{5} = 1 \\ s(\{b,d,e\}) &= \frac{4}{5} = 0.8 \end{aligned}$$

(d) Answer:

$$\begin{aligned} c(bd \longrightarrow e) &= \frac{0.8}{1} = 80\% \\ c(e \longrightarrow bd) &= \frac{0.8}{0.8} = 100\% \end{aligned}$$

(e) There are no apparent relationships between s1, s2, c1, and c2.

Question 4. (15 points)

(a)
$\{1, 2, 3, 4\}$,$\{1, 2, 3, 5\}$,$\{1, 2, 3, 6\}$.
$\{1, 2, 4, 5\}$,$\{1, 2, 4, 6\}$,$\{1, 2, 5, 6\}$.
$\{1, 3, 4, 5\}$,$\{1, 3, 4, 6\}$,$\{2, 3, 4, 5\}$.
$\{2, 3, 4, 6\}$,$\{2, 3, 5, 6\}$.

(b) $\{1, 2, 3, 4\}$, $\{1, 2, 3, 5\}$, $\{1, 2, 4, 5\}$, $\{2, 3, 4, 5\}$, $\{2, 3, 4, 6\}$.

(c) $\{1, 2, 3, 4\}$

Question 5. (20 points)

(a)

	c	$\overline{c}$
b	3	4
$\overline{b}$	2	1

	d	$\overline{d}$
a	4	1
$\overline{a}$	5	0

	d	$\overline{d}$
b	6	1
$\overline{b}$	3	0

	c	$\overline{c}$
e	2	4
$\overline{e}$	3	1

	a	$\overline{a}$
c	2	3
$\overline{c}$	3	2

(b)

(i)

Rules	Support	Rank
$b \longrightarrow c$	0.3	3
$a \longrightarrow d$	0.4	2
$b \longrightarrow d$	0.6	1
$e \longrightarrow c$	0.2	4
$c \longrightarrow a$	0.2	4

(ii)

Rules	Confidence	Rank
$b \longrightarrow c$	$3/7$	3
$a \longrightarrow d$	$4/5$	2
$b \longrightarrow d$	$6/7$	1
$e \longrightarrow c$	$2/6$	5
$c \longrightarrow a$	$2/5$	4

(iii)

Rules	Interest	Rank
$b \longrightarrow c$	0.214	3
$a \longrightarrow d$	0.72	2
$b \longrightarrow d$	0.771	1
$e \longrightarrow c$	0.167	5
$c \longrightarrow a$	0.2	4

(iv)

Rules	IS	Rank
$b \longrightarrow c$	0.507	3
$a \longrightarrow d$	0.596	2
$b \longrightarrow d$	0.756	1
$e \longrightarrow c$	0.365	5
$c \longrightarrow a$	0.4	4

Question 6. (5 points)

$$P(Sam|am) = \frac{1}{7}$$

Question 7. (20 points)

(a) Number of Outputs: n outputs, one per input word at each time step.

Each ˆy (t) is a probability distribution over 4 NER categories.

Each word in the sentence is fed into the RNN and one output is produced at every time step corresponding to the predicted tag/category for each word.

(b) Number of Outputs: 1 output. n outputs is also acceptable if they say for instance to take average of all outputs.

Each ˆy (t) is a probability distribution over 5 sentiment values.

Each word in the sentence is fed into the RNN and one output is produced from the hidden states (by either taking only the final, max, or mean across all states) corresponding to the sentiment value of the sentence.

(c) Number of Outputs: arbitrary

Each ˆy (t) is a probability distribution over the vocabulary.

The previous output is fed as input for the next time step and produces the next output corresponding to the next predicted word of the generated sentence.
As a detail, the first input can also be a designated < ST ART > token and the first output would be the first word of the sentence.