HW0

section1

(1) b d   ① b : $Y \times Y$ ② D : $Y \times Z$

(2) (a) $P(Y=1) = P(X=1,Y=1) + P(X=2,Y=1) + P(X=3,Y=1) = 0.4$

   $P(Y=2) = P(X=1,Y=2) + P(X=2,Y=2) + P(X=3,Y=2) = 0.6$

   $P(X=1) = P(X=1,Y=1) + P(X=1,Y=2) = 0.3$

   $P(X=2) = P(X=2,Y=1) + P(X=2,Y=2) = 0.4$

   $P(X=3) = P(X=3,Y=1) + P(X=3,Y=2) = 0.3$

   $E(X) = 1P(X=1) + 2P(X=2) + 3P(X=3) = 0.3 + 0.8 + 0.9 = 2$

  (b) $E(Y) = 1P(Y=1) + 2P(Y=2) = 0.4 + 1.2 = 1.6$

   $E(X+Y) = E(X) + E(Y) = 3.6$

(3) (a) $P(\text{draw ball is Red}) = \frac{1}{2}P(\text{Red from A}) + \frac{1}{2}P(\text{Red from B})$

   $= \frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{3}{10} = \frac{1}{5} + \frac{3}{20} = \frac{7}{20}$

  (b) $P(\text{ball from A} \mid \text{draw ball is Red}) = \dfrac{\frac{1}{2 \times 5}}{\frac{7}{20}} = \frac{1}{5} \times \frac{20}{7} = \frac{4}{7}$

(4) (a) $f(x) = \frac{1}{1+e^{-x}}$ $1 - f(-x) = 1 - \frac{1}{1+e^{x}} = \frac{e^{x}}{1+e^{x}} = \frac{1}{1+e^{-x}} = f(x)$

   Hence : $f(x) + f(-x) = 1$

  (b) $\frac{\partial f}{\partial x} = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \frac{d}{dx}\left((1+e^{-x})^{-1}\right) = -1 \cdot (1+e^{-x})^{-2} \cdot (-e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^{2}} = f(x)(1-f(x))$

HW1

Question 1 N-Gram Language Models

part (a) $P(W_i | W_i) = \frac{count(W_i, W_{i+1})}{count(W_i)}$

① count(Wi):

the	cat	dog	sat	on	is	mat	floor	\<S\>	\</S\>
5	1	1	2	2	1	2	1	3	3

② count (Wi, Wi+1):

	the	cat	dog	sat	on	is	mat	floor	\<S\>	\</S\>
the	0	1	1	0	0	0	2	1	0	0
cat	0	0	0	1	0	0	0	0	0	0
dog	0	0	0	1	0	0	0	0	0	0
sat	0	0	0	0	1	0	0	0	0	1
on	2	0	0	0	0	0	0	0	0	0
is	0	0	0	0	1	0	0	0	0	0
mat	0	0	0	0	0	1	0	0	0	1
floor	0	0	0	0	0	0	0	0	0	1
\<S\>	3	0	0	0	0	0	0	0	0	0
\</S\>	0	0	0	0	0	0	0	0	0	0

③ $P(the | \<S\>) = \frac{3}{3} = 1$

$P(cat | the) = \frac{1}{5}$ $P(dog | the) = \frac{1}{5}$ $P(mat | the) = \frac{2}{5}$ $P(floor | the) = \frac{1}{5}$

$P(sat | cat) = \frac{1}{1} = 1$

$P(on | sat) = \frac{1}{2}$ $P(\</S\> | sat) = \frac{1}{2}$

$P(the | on) = \frac{2}{2} = 1$

$P(\</S\> | floor) = \frac{1}{1} = 1$

$P(sat | dog) = \frac{1}{1} = 1$

$P(on | is) = 1$

$P(is | mat) = \frac{1}{2}$ $P(\</S\> | mat) = \frac{1}{2}$

part (b) $P(w_1 \cdots w_n) = \prod P(w_i | w_{i-1})$ if $P(w_i | w_{i-1}) = 0$, then $P(w_1 \cdots w_n) = 0$

    Example 1: ⟨S⟩ the cat is on the mat ⟨/S⟩

              since $P(is | cat) = 0$, the sentence has zero prob

    Example 2: ⟨S⟩ the dog is on the floor

              since $P(is | dog) = 0$, the sentence has zero prob

part (c) $P(\text{"⟨S⟩ the cat sat on the mat ⟨/S⟩"}) = 1 \times \frac{1}{5} \times 1 \times \frac{1}{2} \times 1 \times \frac{2}{5} \times \frac{1}{2} = \frac{1}{50}$

    $P(\text{"⟨S⟩ the cat sat ⟨/S⟩"}) = 1 \times \frac{1}{5} \times 1 \times \frac{1}{2} = \frac{1}{10}$

    $\therefore \frac{1}{10} > \frac{1}{50}$

    The statement is false

part (d) prefix = "⟨S⟩ the mat is on the"

    $P(cat | the) = \frac{1}{5}$ $P(dog | the) = \frac{1}{5}$ $P(mat | the) = \frac{2}{5}$ $P(floor | the) = \frac{1}{5}$

    ① let $W_6 = \text{"mat"}$, then $P(is | mat) = \frac{1}{2}$ $P(⟨/S⟩ | mat) = \frac{1}{2}$

        let $W_7 = \text{"⟨/S⟩"}$, then "⟨S⟩ the mat is on the mat ⟨/S⟩"

        and its prob should be $P(mat | the) \times P(⟨/S⟩ | mat) = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$

    ② let $W_6 = \text{"floor"}$, then $P(⟨/S⟩ | floor) = 1$

        let $W_7 = \text{"⟨/S⟩"}$, then "⟨S⟩ the mat is on the floor ⟨/S⟩"

        and its prob should be $P(⟨/S⟩ | floor) \, P(floor | the) = \frac{1}{5} \times 1 = \frac{1}{5}$

part(e)   the cat sat on the $\left\{\begin{array}{l} \text{dog} \to \text{sat} \left\{\boxed{\text{on}} \to \text{the} \atop \text{<IS>} \right. \\ \text{is} \to \boxed{\text{on}} \to \text{the} \\ \text{mat} \left\{\text{<IS>}\right. \\ \text{floor} \to \text{<IS>} \end{array}\right.$

we can see from the graph that it has 2 cycles, hence the number of the total number of distinct non-zero probability sentences could be infinite. But most of the sentences are pretty strange (e.g: the cat sat on the mat is on the cat ...)

   N-gram model would help by incorporating more context and reducing such repetitions, though it may suffer from data sparsity

① Example for Bigram: the cat sat on the dog sat on the → ...

                    the cat sat on the mat is on the → ...

② Example for Trigram: the cat sat on the $\left\{\begin{array}{l} \text{mat} \to \text{is on the} \to ... \\ \text{floor} \to \text{<IS>} \end{array}\right.$

   Hence, in trigram, some of the cycles has been broken

Question 3 Word Embeddings

part(a) positive pairs: (ball, beach), (ball, dog), (cafe, cat), (cat, dog)

negative pairs: (park, cafe), (ball, cat), (cat, ball)

① update times: i) target: ball: 3

cafe: 1

cat: 2

park: 1

ii) context: beach: 1

dog: 2

cat: 2

ball: 1

cafe: 1

② appear in both negative and positive pairs

(i) (ball, beach, +), (ball, dog, +), (ball, cat, −)

(ii) (cat, dog, +), (cat, ball, −), (cafe, cat, +), (ball, cat, −)

(iii) (cafe, cat, +), (park, cafe, −)

Hence: "ball" and "cat" change most cause they appear most in the target word, and they appear in both positive and negative pairs

part(b) (w=cafe, c=cat, y=+) ⇒ let wcafe and ccat be similar
            (w=ball, c=cat, y=-) ⇒ let wcafe and ccat be unrelated
      Therefore, the model is expected to learn that cat is related to cafe
but unrelated to ball