Chapter 6

Ex 1

$$z_{pos} = \sigma(w, c_{pos}) = \frac{1}{1 + e^{-c_{pos} w}}$$

$$z_{neg} = 1 - \sigma(w, c_{neg})$$

$$L = -\log z_{pos} - \sum_{i=1}^{k} \log z_{neg}$$

* $\frac{dL}{dz_{pos}} = -\frac{1}{z_{pos}}$; * $\frac{dz_{pos}}{dc_{pos}} = w(z_{pos} - z_{pos}^2)$

* $\frac{dL}{dz_{neg}} = -\frac{1}{z_{neg}}$; * $\frac{dz_{neg}}{dc_{pos}} = 0$

* Applying chain rules:

$$\frac{dL}{dc_{pos}} = \frac{dL}{dz_{pos}} \cdot \frac{dz_{pos}}{dc_{pos}} + \frac{dL}{dz_{neg}} \cdot \frac{dz_{neg}}{dc_{pos}} = (z_{pos} - 1) \cdot w = (\sigma(c_{pos} \cdot w) - 1) w$$

equation 6.35: $\boxed{\frac{dL}{dc_{pos}} = (\sigma(c_{pos} \cdot w) - 1) w}$

* $\frac{dz_{pos}}{dc_{neg}} = 0$; * $\frac{dz_{neg}}{dc_{neg}} = w(\sigma^2(c_{neg} \cdot w) - \sigma(c_{neg} \cdot w))$

* Applying chain rules:

$$\frac{dL}{dc_{neg}} = \frac{dL}{dz_{pos}} \cdot \frac{dz_{pos}}{dc_{neg}} + \frac{dL}{dz_{neg}} \cdot \frac{dz_{neg}}{dc_{neg}} = \frac{1}{\sigma(c_{neg} \cdot w) - 1} \cdot w(\sigma^2(c_{neg} \cdot w) - \sigma(c_{neg} \cdot w))$$

$$= w \cdot \sigma(c_{neg} \cdot w)$$

Equation 6.36: $\boxed{\frac{dL}{dc_{neg}} = w \cdot \sigma(c_{neg} \cdot w)}$

* $\frac{dz_{pos}}{dw} = c_{pos} \cdot \left(\sigma(c_{pos} \cdot w) - \sigma^2(c_{pos} \cdot w) \right)$

* $\frac{dz_{neg}}{dw} = c_{neg} \cdot \left(\sigma^2(c_{neg} \cdot w) - \sigma(c_{neg} \cdot w) \right)$

* Applying chainrule

$$\frac{dL}{dw} = \frac{dL}{dz_{pos}} \cdot \frac{dz_{pos}}{dw} + \frac{dL}{dz_{neg}} \cdot \frac{dz_{neg}}{dw} = c_{pos} \cdot \left(\sigma(c_{pos} \cdot w) - 1 \right) + \sum_{i=1}^{k} c_{neg} \cdot \sigma(c_{neg} \cdot w)$$

Equation 6.37: $\boxed{\frac{dL}{dw} = c_{pos} \left[\sigma(c_{pos} \cdot w) - 1 \right] + \sum_{i=1}^{k} c_{neg}^{i} \, \sigma(c_{neg}^{i} \cdot w)}$

Ex 2.

context \ word	dogs	bark	loudly	cats	meow	softly	and	play
dogs	0	1	0	0	0	0	1	0
bark	1	0	1	0	0	0	0	0
loudly	0	1	0	0	0	0	0	0
cats	0	0	0	0	1	0	1	1
meow	0	0	0	1	0	1	0	0
softly	0	0	0	0	1	0	0	0
and	1	0	0	1	0	0	0	0
play	0	0	0	1	0	0	0	0

$$P(w,c) = \frac{f_{wc}}{\sum_{i \in W} \sum_{j \in C} f_{ij}} \quad ; \quad P(w) = \frac{\sum_{j \in C} f_{wj}}{\sum_{i \in W} \sum_{j \in C} f_{ij}} \quad ; \quad P(c) = \frac{\sum_{i \in W} f_{ic}}{\sum_{i \in W} \sum_{j \in C} f_{ij}}$$

(dogs, bark): $PMI = \log_2 \frac{P(dogs, bark)}{P(dogs) \cdot P(bark)} = \log_2 \frac{\frac{1}{13}}{\frac{2}{13} \cdot \frac{2}{13}} = 1,7$

$PPMI = \max(PMI, 0) = 1,7$

(cats, meow) $PMI = \log_2 \frac{P(cats, meow)}{P(cats) \cdot P(meow)} = \log_2 \frac{13}{6} = 1,11$

$PPMI = \max(PMI, 0) = 1,11$

(dogs, cats) $PMI = \log_2 \frac{P(dogs, cats)}{P(dogs) \cdot P(cats)} = \log_2 0$

$PPMI = max(PMI, 0) = 0$

Ex 3

$$\cos(king, queen) = \frac{0{,}7 \cdot 0{,}69 + 0{,}1 \cdot 0{,}12 + 0{,}3 \cdot 0{,}31}{\sqrt{0{,}7^2 + 0{,}1^2 + 0{,}3^2} \cdot \sqrt{0{,}69^2 + 0{,}12^2 + 0{,}31^2}} = 0{,}999$$

$$\cos(king, man) = \frac{0{,}7 \cdot 0{,}5 + 0{,}1 \cdot 0{,}09 + 0{,}3 \cdot 0{,}4}{\sqrt{0{,}7^2 + 0{,}1^2 + 0{,}3^2} \cdot \sqrt{0{,}5^2 + 0{,}09^2 + 0{,}4^2}} = 0{,}964$$

King is nearer queen than man for some reasons:

- queen is more likely appeared in the same context of king than man. in the training documents.

- The word "man" is used so frequently in many contexts, which isn't tied in the contexts of "king", so Its cosine similarity is lower than queen.

Ex 4

1.

Equation for TF-IDF:

$$* \ TF\text{-}IDF(t, d) = \log \frac{count(t, d)}{\sum_{w \in d} count(w, d)} \cdot \log \frac{N}{1 + \{d \in D \text{ and } t \in d\}}$$

TF-IDF is usually used for representing term-document vector

Equation for PPMI:

$$* \ PPMI(w, c) = max\left(\log \frac{P(w, c)}{P(w) \cdot P(c)}, 0\right)$$

+ $P(w, c)$: Probabilities that c in the window context of w

$$P(w, c) = \frac{f(w, c)}{\sum_{i \in w} \sum_{j \in c} f(i, j)}$$

+ $P(w)$ Probabilities that w appeared in the window context

$$P(w) = \frac{\sum_{j \in c} f_{w,j}}{\sum_{i \in w} \sum_{j \in c} f(i,j)}$$

+ P(C) Probabilities that w appeared in the window context

$$P(c) = \frac{\sum_{i \in w} f_{ic}}{\sum_{i \in w} \sum_{j \in c} f(i,j)}$$

PPMI is usually used for representing term-term vector

2.

- Words that rarely appeared in the training documents are assigned with high weights due to the sparsity of the data. This problem can significantly bias the predictions
- PPMI cannot used to represent out-of-vocabulary words

3.

- Advantages:
  + Easier to compute, understand
  + Fast to compute
  + Require smaller spaces.
- Disadvantages
  + Cannot represent well the meaning of the words
  + Cannot capture the contexts.
  + Create sparse vectors, which cause unefficiency.
  + out-of-vocabulary words cannot be represented.