

# PA1 Report

Kevin Lin

1/30/2025

## Part 3: Understanding Skip-Gram

### Q1

3a) We are given the skip-gram model defined as:

$$P(\text{context} = y | \text{word} = x) = \frac{\exp(\mathbf{v}_x \cdot \mathbf{c}_y)}{\sum_{y'} \exp(\mathbf{v}_x \cdot \mathbf{c}_{y'})}$$

where  $x$  is the "center word",  $y$  is a "context word" being predicted, and  $\mathbf{v}_x$  and  $\mathbf{c}_y$  are  $d$ -dimensional vectors corresponding to words and contexts respectively. Each word has independent vectors for each, thus each word has two embeddings.

Given the sentences:

the dog  
the cat  
a dog

window size of  $k = 1$ , we get the training examples: ( $x = \text{the}, y = \text{dog}$ ), ( $x = \text{dog}, y = \text{the}$ ). Consequently, the skip-gram objective, log-likelihood is  $\sum_{(x,y)} \log P(y|x)$ . With word and context embeddings of dimension  $d = 2$ , the context embedding vectors  $w$  for *dog* and *cat* are both  $(0, 1)$ , and the embeddings vectors  $w$  for *a* and *the* are  $(1, 0)$ . Thus, the set of probabilities  $P(y|\text{the})$  that maximize the log-likelihood are:

$$P(y|\text{the}) = \begin{cases} \frac{1}{2} & y = \text{dog} \\ \frac{1}{2} & y = \text{cat} \\ 0 & y = \text{a} \\ 0 & y = \text{the} \end{cases}$$

3b) We want a setting  $\mathbf{v}_{\text{the}}$  where  $P(\text{dog}|\text{the}) \approx 0.5$ ,  $P(\text{cat}|\text{the}) \approx 0.5$ , and  $P(\text{a}|\text{the}), P(\text{the}|\text{the}) \approx 0$ .