e. (3 points) In your second attempt, you are going to instead model each bigram as a sum of the two words. That is, instead of having a vector for b(cat, saw), we instead represent this as $v_{\text{cat}} + v_{\text{saw}}$. Contexts are still unigrams.

Modify the skip-gram formula accordingly and write the formula below, introducing notation as needed.

word = s (sum of two unigrams)

$$p(\text{context}=y \mid \text{word}=s) = \frac{\exp(v_s^T c_y)}{\sum_{y' \in \text{vocab}} \exp(v_s^T c_{y'})}$$

f. (2 points) What is the big-O runtime of computing the probability for a single bigram-context pair under this new scheme? Express this in terms of the quantities in part (b).

$O(Vd)$

g. (2 points) How many parameters are in the model? Express this in terms of the quantities in part (b).

$2 \cdot V \cdot d$