Q2:

The key difference between reinforcement learning and supervised learning is whether having a labeled target.

In supervised learning, the learning process is driven by a labeled dataset consisting of input-output pairs. The model is trained to map inputs to the correct output using these examples.

In reinforcement learning does not require labeled input-output pairs. Instead, it learns from interactions with an environment, which are structured as states, actions, and rewards.

In reinforcement learning, whether an agent receives a reward immediately after taking an action is not guaranteed and depends on the nature of the task and environment:

Immediate Rewards: In some environments, the consequences of an action and thus the rewards can be immediate. For example, in a game where points are scored by hitting a target, the reward (points) is received right after the action (throwing or shooting).

Delayed Rewards: In many cases, especially in complex environments, rewards are delayed. This means the full impact of an action may not be apparent until much later. For instance, strategic decisions in a game like chess may have consequences that only become clear several moves later.

The structure of rewards (immediate vs. delayed) is crucial because it influences how the learning algorithm plans and evaluates actions. Immediate rewards can simplify learning by providing quick feedback, while delayed rewards require the algorithm to consider longer-term consequences and possibly develop more sophisticated strategies.

Q3:

1).

The minimal number of vector dot product should be 100. Because if we can store the result of each dot product, when we are calculating the softmax part, we will get the dot product of all 100 words' vectors' with the current word.

The most expensive part should also be the calculating the softmax part as it will calculate 100 dot vectors and it involves exponentiations, summations, and divisions.

To solve this, we can use negative sampling to reduce the cost by randomly samples a small number of words which are not in context.

2).

Here should be 26 vector dot products to calculate the loss. The function can be divided into two parts. The first part is to calculate the dot products of current word and target context word. The second part is to calculate the sum of dot products of negative sampling and current words. Since K = 25, we need to calculate 25 time in the second part.