Q1

To use negative sampling, we first create a distribution that is used to select negative samples. The noise distribution is typically a unigram distribution, which means that the probability of selecting a word is proportional to the frequency of the word in the training corpus.
Once we have created a noise distribution, we can use it to select negative samples. For each training example, we select a positive sample and a number of negative samples. The positive sample is the word that is in the context of the current word. The negative samples are words that are not in the context of the current word.
We then modify the loss function so that we maximise the loss or minimise the negative similarity between different words and maximise the similarity between similar words.
The weights of the neurons that are associated with the positive sample are updated to increase the probability of the positive sample. The weights of the neurons that are associated with the negative samples are updated to decrease the probability of the negative samples.

Q2

Semantic similarity is the measure of how similar two words are in meaning.
Word embeddings are vector representations of words that capture their semantic meaning. Once word embeddings are learned, they can be used to measure the semantic similarity of words.

There are a variety of techniques for measuring semantic similarity using word embeddings. Two of the most common techniques are:

Cosine similarity: Cosine similarity is a measure of the cosine of the angle between two vectors. In the case of word embeddings, the vectors represent the meaning of the words. Therefore, the cosine similarity between two words is a measure of how similar their meanings are.
Euclidean distance: Euclidean distance is a measure of the distance between two points in a Euclidean space. In the case of word embeddings, the points represent the meaning of the words. Therefore, the Euclidean distance between two words is a measure of how different their meanings are.