**Problem 1: 1D CNN for Text Classification**

* Smaller filters capture local patterns like short phrases or n-grams. Medium filters capture slightly longer patterns. Using multiple filter sizes allows the model to capture patterns at different scales. In the results, the model with multiple filter sizes typically performs better because it can learn both short and medium-range patterns in the text. However, more filters increase computational cost and may lead to overfitting if not properly regularized.

**Problem 2: POS Tagging using Viterbi HMM**

1. Accuracy:
   1. The Maximum Entropy tagger typically achieves higher accuracy than the HMM Viterbi tagger because it can incorporate more contextual features beyond just the previous tag
2. Generalization:
   1. Maximum Entropy models generalize better to unseen words because they can use morphological features
3. Computational Efficiency:
   1. HMM with Viterbi is generally faster to train and decode, especially with larger tag sets.
   2. Maximum Entropy models require more computational resources for training but can be efficient during prediction

**Problem 3: RNN-based Text Generation**

1. Vanilla RNN:
   1. Simplest architecture, often produces less coherent text
   2. Tends to repeat common patterns
   3. May struggle with long-term dependencies
2. Stacked RNN:
   1. Deeper architecture captures more complex patterns
   2. Generally, produces more coherent text than vanilla
   3. Better at maintaining context over longer sequences
   4. Slower to train due to increased complexity
3. Bidirectional RNN:
   1. Captures context from both past and future
   2. Often produces the most grammatically correct text
   3. May sometimes lose character “voice” as it averages context
   4. Most computationally intensive
4. The stacked and bidirectional RNNs typically produce higher quality text with better coherence and grammar, while the vanilla RNN is faster to train but produces less impressive results.