Exercise: POS Tagging with Hidden Markov Models

PART 1 — Training Data & HMM Learning

1.1 Labeled Corpus

We are given the following small labeled corpus. Each word is paired with its Part-of-Speech tag.
Tag set: NN (Noun), VB (Verb), DT (Determiner)

Sentence 1:

```
the/DT  dog/NN  runs/VB
```

Sentence 2:

```
the/DT  lion/NN  runs/VB
```

Sentence 3:

```
dog/NN  is/VB an/DT animal/NN
```

Sentence 4:

```
cat/NN  is/VB an/DT  animal/NN
```

1.2 Learning Task

Using the corpus above, compute the following components of the HMM:

1. Q
2. Vocabulary
3. Prior (Initial) Probabilities π(tag)
4. Transition Matrix A
5. Emission Matrix B

```
Hint: π(tag) = count(tag starts sentence) / total sentences
```

PART 2 — Viterbi Decoding

2.1 Task

Given the HMM you computed in Part 1, use the Viterbi algorithm to find the most likely tag sequence for the following sentence:

```
"the   cat   runs"
```