Q1

Implement the latent Dirichlet allocation (LDA) model to generate a corpus from a given set of parameters. Build a function `lda()` that takes four arguments:

1. `vocabulary` - list (of length V) of strings
2. `beta` - topic-word matrix, numpy array of size (k, V)
3. `alpha` - topic distribution parameter vector, of length k
4. `xi` - Poisson parameter (scalar) for document size distribution

and returns:

1. `w` - list of words (strings) in a document

Demonstrate using this function with the following parameters:

```
vocabulary = ['bass', 'pike', 'deep', 'tuba', 'horn', 'catapult']
beta = np.array([
    [0.4, 0.4, 0.2, 0.0, 0.0, 0.0],
    [0.0, 0.3, 0.1, 0.0, 0.3, 0.3],
    [0.3, 0.0, 0.2, 0.3, 0.2, 0.0]
])
alpha = np.array([1, 3, 8])
xi = 50
```

Q2

Generate a corpus of documents using the same parameters and use an LDA solver e.g. `gensim` or pip's `lda` to attempt to infer the parameters.

Submit your solution as a Jupyter notebook (.ipynb file).