Question:

Take the simple sentences we used in our word example from the lecture.

Put these into the program and compute the K-L divergence scores for them, in both directions.

The simple sentences:

```
d1 = """Many research publications want you to use BibTeX, which better
organizes the whole process. Suppose for concreteness your source
file is x.tex. Basically, you create a file x.bib containing the
bibliography, and run bibtex on that file."""

d2 = """In this case you must supply both a \left and a \right because the
delimiter height are made to match whatever is contained between the two commands.
But, the \left doesn't have to be an actual 'left
delimiter', that is you can use '\left)' if there were some reason
to do it."""

d3 = """Many research publications want you to use BibTeX, which better
organizes the whole process. Suppose for concreteness your source
file is x.tex.But, the \left doesn't have to be an actual 'left
delimiter', that is you can use '\left)' if there were some reason
to do it."""
```

The K-L divergence scores of them:

```
KL-divergence between d1 and d2: 6.52185430963571
KL-divergence between d2 and d1: 6.511423630945803
KL-divergence between d2 and d3: 3.005348407337924
KL-divergence between d3 and d2: 3.2024640237568587
KL-divergence between d1 and d3: 1.872591587570143
KL-divergence between d3 and d1: 3.213031026043869
```

Now create a third story that is very different to the other two, add it to the program and report how its score changes relative to the first two.

the third story:

```
d4="""Take the simple sentences we used in our word example from the lecture.
Put these into the program and compute the K-L divergence scores for them, in both directions.
Now create a third story that is very different to the other two,
add it to the program and report how its score changes relative to the first two.
"""
```

the K-L divergence scores:

```
KL-divergence between d1 and d4: 7.137401069984823
KL-divergence between d4 and d1: 6.794727590664626
KL-divergence between d2 and d4: 6.8378380990555545
KL-divergence between d4 and d2: 6.4076058872588275
```

Comment on whether the scores make sense.

Explain what role epsilon and gamma play in the computation of K-L.

The formula of K-L divergence score is :

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right).$$

The smaller the relative entropy, the more similar the distribution of the two variables, and conversely.

Among story d1, d2, and d3:

From the K-L divergence score we got, we can see that , the KL-divergence between d1 and d2 is 6.52185430963571 which is the maximum value, it is because that there are too many differences between Document 1 and Document 2. On the other hand, KL-divergence between d1 and d3 is 1.872591587570143 which is the minimum value, and the K-L divergence between d2 and d3 is 3.005348407337924 which is the second smallest value, we know that there is not much difference between d1 and d3, also d2 and d3.

Among story d1, d2, and d4:

Because d4 is totally different from d1 and d2, so KL-divergence between d1 and d4 is 7.137401069984823 which is even bigger than the KL-divergence between d1 and d2, and the KL-divergence between d4 and d1: 6.794727590664626, KL-divergence between d2 and d4: 6.8378380990555545, KL-divergence between d4 and d2: 6.4076058872588275. All of them are big numbers.

Furthermore Because that the measure is asymmetrical, for instance, the KL-divergence between d1 and d3 and KL-divergence between d3 and d1 is not the same, so it is not strictly a distance metric.

Explain what role epsilon and gamma play in the computation of K-L.

Epsilon is the probability of a term which is not in a document. It is set to a small value instead of 0 to avoid the distance to be infinite. Gamma is a normalization coefficient to account of epsilon, so a probability of a term in a category satisfies the properties of a probability (sum to 1).