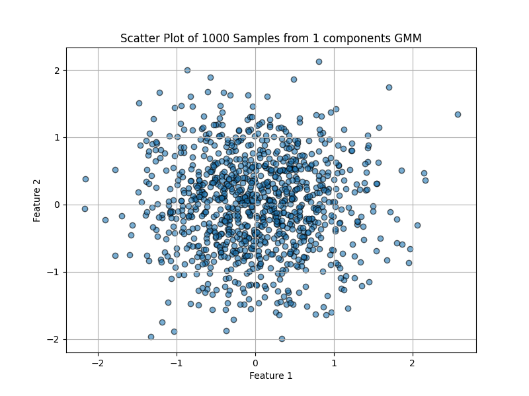
**IML Project 5 Report - Nir Ellor**

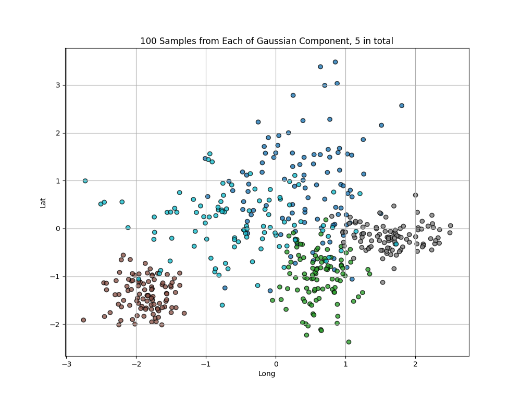
**GMM:**

1. Following are the plots for all 1, 5, 10 and 33 components, as on the left side we sample 1000 samples from the GMM and on the right side we sample 100 samples from each Gaussian:

1.a: 1.b:

תמונה שמכילה טקסט, תרשים, קו, עלילה

התיאור נוצר באופן אוטומטי****



תמונה שמכילה טקסט, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, מפה, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים

התיאור נוצר באופן אוטומטי

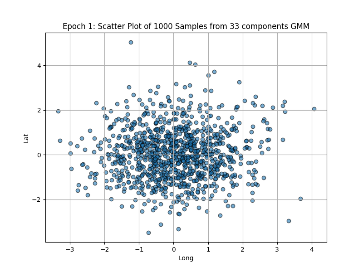
The scatter plots demonstrate that as the number of GMM components increases, the synthetic samples better capture the complexity and structure of the underlying data distribution.

2.

2. a:

Following are plots with random initialization for all 1, 10, 20, 30, 40 and 50 epochs:

תמונה שמכילה צילום מסך, תרשים, צבעוני, טקסט

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, מפה

התיאור נוצר באופן אוטומטיתמונה שמכילה תרשים, טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, מפה, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה תרשים, טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, צבעוני

התיאור נוצר באופן אוטומטיתמונה שמכילה צילום מסך, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, צילום מסך, צבעוני, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, מפה, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים

התיאור נוצר באופן אוטומטי

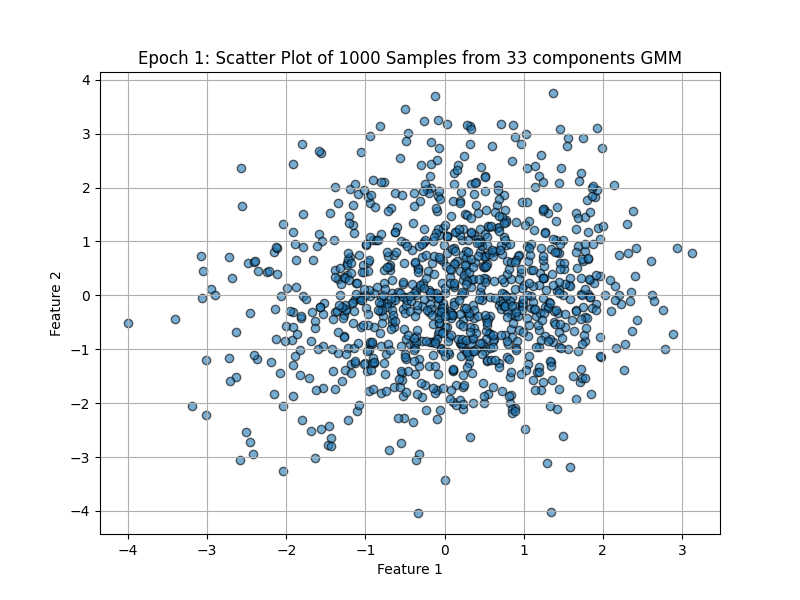
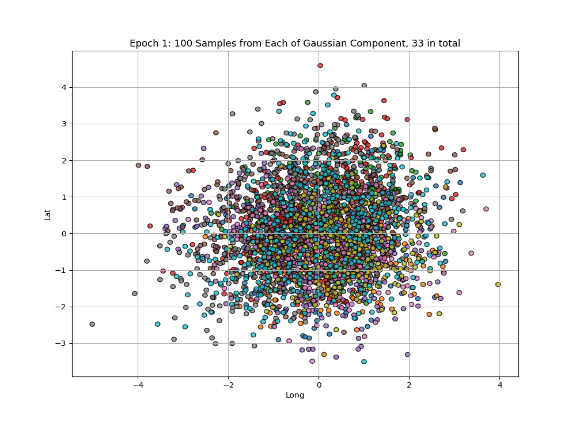
תמונה שמכילה טקסט, קו, עלילה, תרשים

התיאור נוצר באופן אוטומטי2. b:

תמונה שמכילה טקסט, קו, עלילה, תרשים

התיאור נוצר באופן אוטומטי2.c:

Following are plots with country means initializations for all 1, 10, 20, 30, 40 and 50 epochs:



תמונה שמכילה טקסט, תרשים, צילום מסך

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, תרשים, מפה

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, תרשים

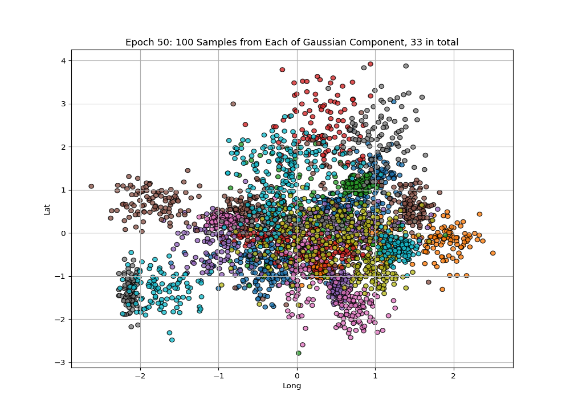
התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, מפה, תרשים

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, צילום מסך, תרשים

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, תרשים

התיאור נוצר באופן אוטומטי

As of 2.a, the plots reveal how the GMM evolves at different epochs when the number of components matches the number of labels (33):

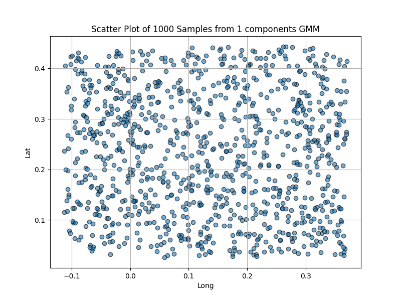
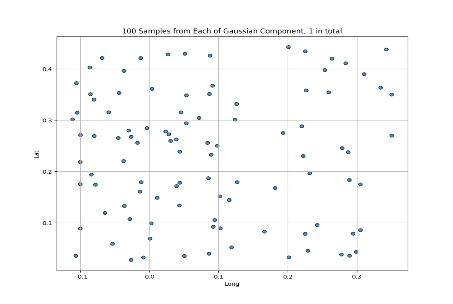
1. **Early Epochs (Epoch 1)**: The model has not yet learned distinct clusters, resulting in scattered and overlapping samples, as seen in both general and conditional plots.
2. **Mid Epochs (Epochs 10–30)**: Clusters become more distinct, and the conditional samples show clearer separation, indicating that the GMM is learning the data structure.
3. **Late Epochs (Epoch 40, 50)**: The clusters stabilize, with well-defined separation in the conditional plots, suggesting that the GMM converged to a meaningful representation of the data.

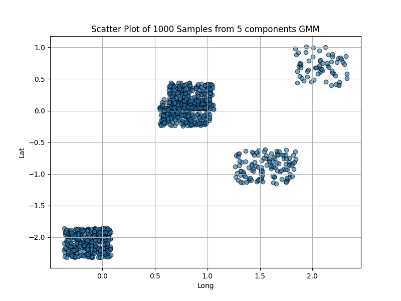
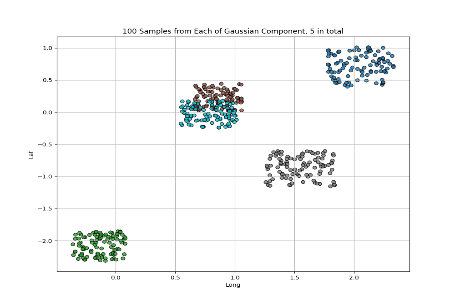
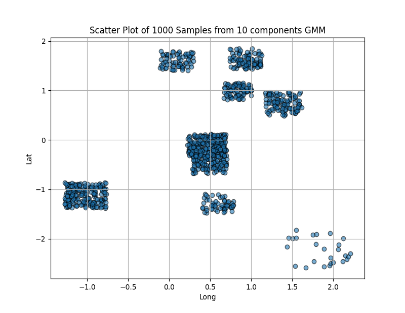
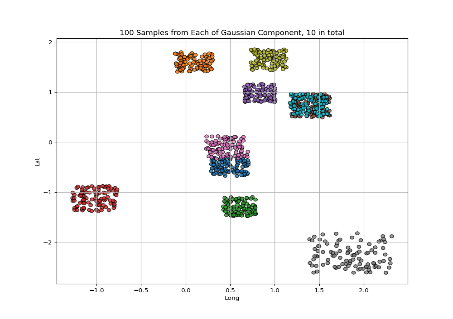
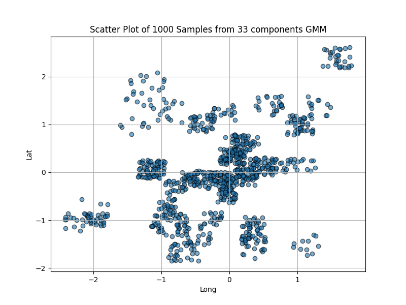
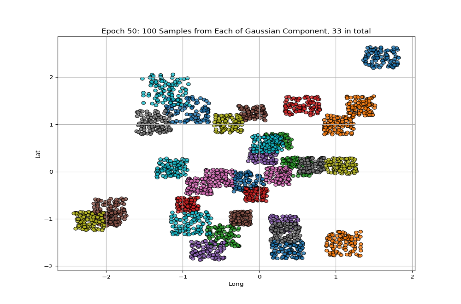
As of 2.b, differences between the plots are minor, with initialization using means showing slightly better performance between epochs 10 and 20. Both approaches achieve similar training and testing log-likelihoods by the final epoch.

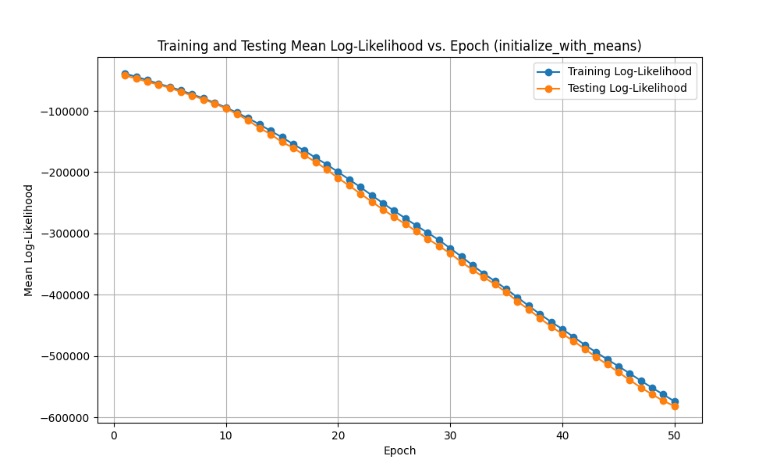
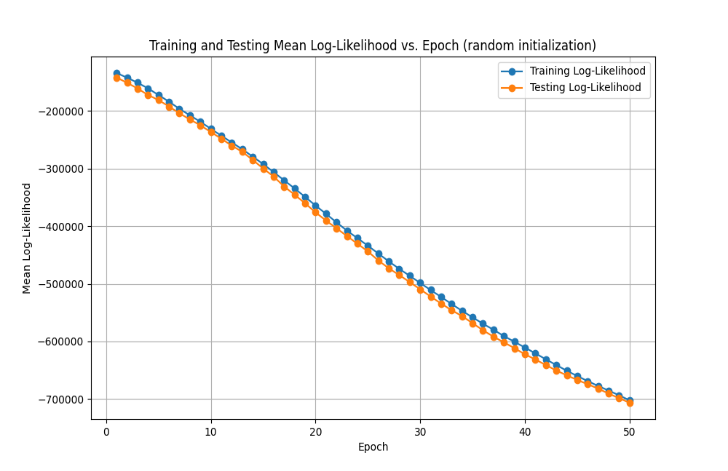
As of 3.b, compared to previous 12 plots, the 12 current plots initialized with means don't appear to have radical differences across epochs 10, 20, 30, 40, and 50. The clusters look reasonably well-formed early on, and their structure stabilizes quickly.

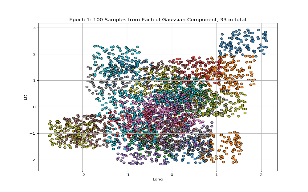
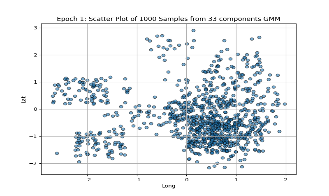
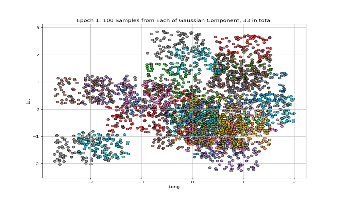
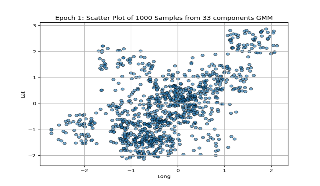
However, initializing with means leads to faster convergence, resulting in well-formed clusters as early as epoch 10, with minimal changes in structure across later epochs.

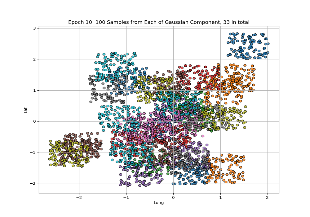
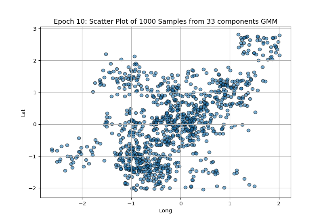
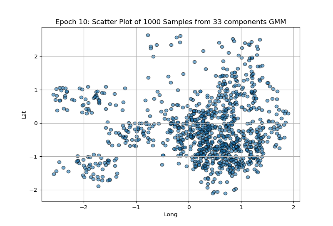
In conclusion, initializing with mean locations accelerates convergence and provides more structured clusters early in training, though final epochs are comparable to random initialization.

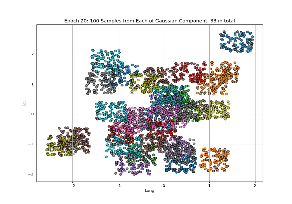
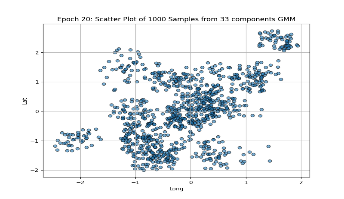
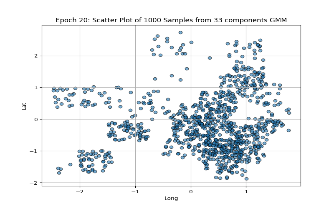
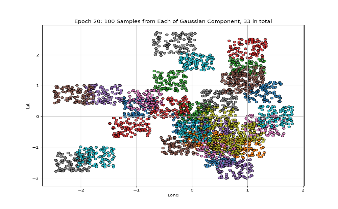
**UMM:**

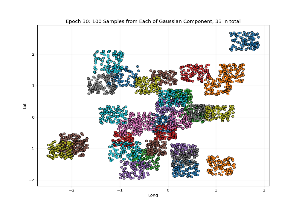
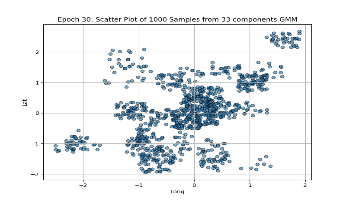
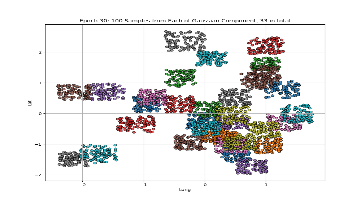
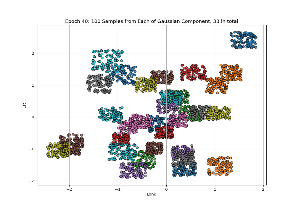
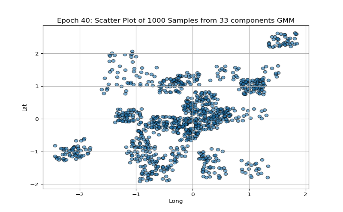


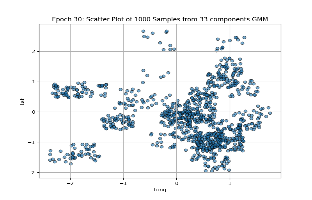
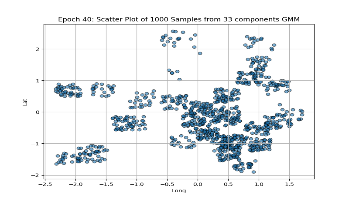
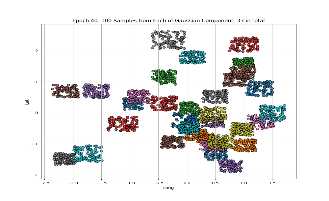
****

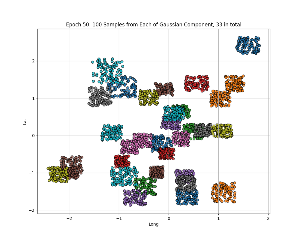
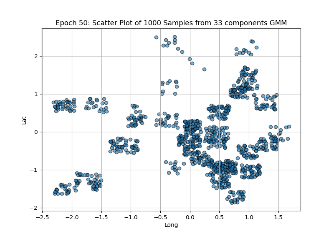
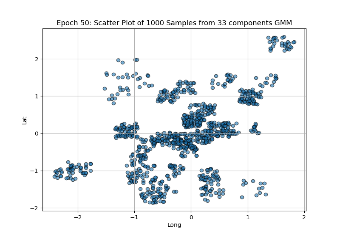
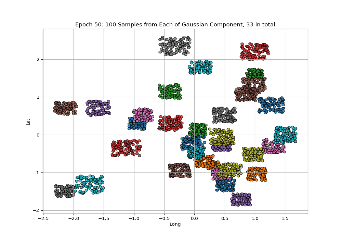










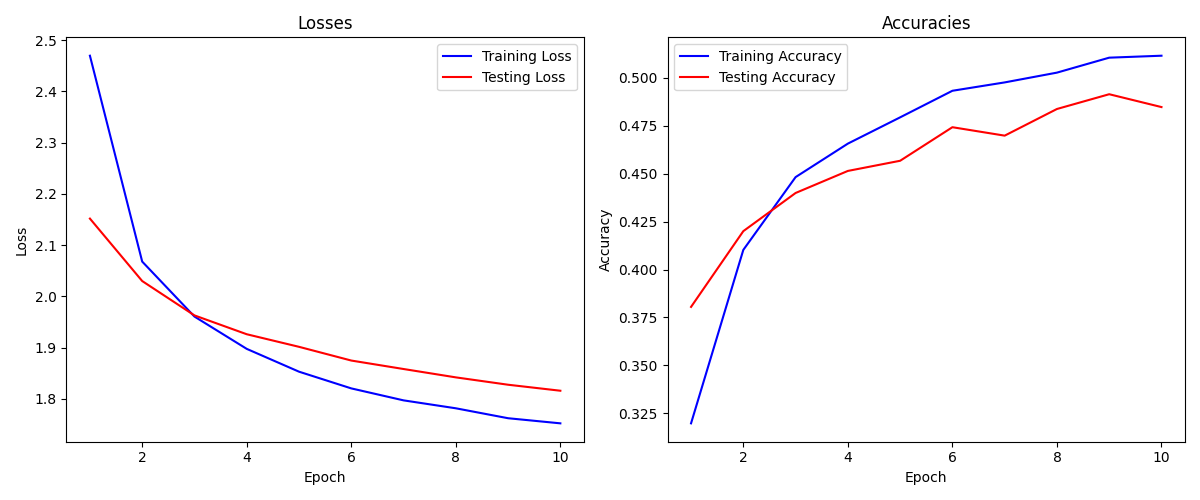


As we advance through the epochs, the supports of the uniform components in the UMM consistently shrink and concentrate around regions of higher data density. The centers of the uniforms progressively align with the centroids of the data clusters, improving separation and reducing overlap between components. This occurs because gradient descent optimizes the parameters to maximize the log-likelihood, which encourages tighter supports around dense areas of the data. However, this process highlights a challenge unique to UMM: the hard boundaries of the uniforms lead to sparse or zero gradients for out-of-bound samples, causing slower or unstable convergence. Unlike the Gaussian Mixture Model (GMM), where smooth Gaussian functions provide non-zero gradients for all data points, the UMM struggles with boundary effects and sparse gradient updates, making optimization more sensitive and less robust.

**Transformer:**

|  |  |  |
| --- | --- | --- |
| **Epochs** | **Without Top k** | **With Top k** |
| **Epoch 1** | | |
| **Sentence 1** | the with crenter:  KI thell ifsream | the a sof seay.    Why the theas,  Al |
| **Sentence 2** | the were a the a-sful, wolse steer | the somen at thers shat as masting |
| **Sentence 3** | the ther cald thifey teer!  And iss | the warth hard se thou trour the s |
| **Epoch 2** | | |
| **Sentence 1** | the Rother siglantion I Sway forli | the of thou have him thank, wolds, |
| **Sentence 2** | the do you leez  qulaine kel.  And I | the here strue werell he all so th |
| **Sentence 3** | the them uveriever uncuncues the g | the way sir,  I'd that a seed thy t |
| **Epoch 3** | | |
| **Sentence 1** | the do will antremishal, thou tear | the too sire wither his hate hoppa |
| **Sentence 2** | the of the begor hard wiFill for t | the hand hims, the will an that ma |
| **Sentence 3** | the ste herear, sus whake be tear | the hasperear hand the she the wic |
| **Epoch 4** | | |
| **Sentence 1** | the shouldour lives love a werear. | the well sir, wast that to too sea |
| **Sentence 2** | the your plerefor city his there s | the as arate think and a mess waye |
| **Sentence 3** | the in ejord of heaver amsty, my a | the the world  If live, sond the th |
| **Epoch 5** | | |
| **Sentence 1** | the cause vile, bettles you would | the some of have me cause, and mos |
| **Sentence 2** | the woils, 'tway make Jortes? Crha | the as beginst the wing inturled m |
| **Sentence 3** | the all word Go, whom see I plalio | the homing we with are more mard t |
| **Epoch 6** | | |
| **Sentence 1** | the lovers with fall, and cue,  For | the breed woe made the crievend al |
| **Sentence 2** | the by hand the make the her freed | the with the dist is any to here s |
| **Sentence 3** | the abu crove, you days best never | the to her the me to shall'd  As ma |
| **Epoch 7** | | |
| **Sentence 1** | the much may I shore una out so de | the hath haste here such and wise |
| **Sentence 2** | the the feal dolt a his depost her | the as and with our shame,  In shap |
| **Sentence 3** | the the felleady.  This itizen.  KI | the a parded midy aliest to his cr |
| **Epoch 8** | | |
| **Sentence 1** | the watch bidgar:  O, ho, than the | the she had stroyed.  CLETESP:  Net |
| **Sentence 2** | the los ame of Poy'd, other:  Tety | the the so hearth.  CLIFORD:  I way |
| **Sentence 3** | the had what and bid your lace.  S | the than him shornow  Shalls to she |
| **Epoch 9** | | |
| **Sentence 1** | the Frannener the be  indot to the | the hold with thee to she before a |
| **Sentence 2** | the upon widly to  Edmise, nato gol | the then soun than hole, and the s |
| **Sentence 3** | the soul back  No holp our thread l | the sorrow out shame  Wer sir, ter |
| **Epoch 10** | | |
| **Sentence 1** | the of forsciusitess, sinse onatua | the the shall no the sun  The son. |
| **Sentence 2** | the sorn?  DUKE OFW:  This ralloye | the with and bother son,  And so hi |
| **Sentence 3** | the thou we suffenter gu'd you sha | the tail.  POLIXENES:  It nature  Th |

Top-k sampling noticeably improves the quality of the generated text compared to the unfiltered approach. Without top-k sampling, the model produces highly random and nonsensical outputs, especially in early epochs, as it samples from the entire vocabulary. In contrast, top-k sampling restricts predictions to the most probable tokens, resulting in more coherent and grammatically structured sentences, even in earlier epochs. However, both methods struggle with producing meaningful, contextually relevant text, likely due to the model's small size, limited training, or dataset constraints. While top-k sampling mitigates randomness, further improvements in model architecture, training duration, and sampling techniques are needed for high-quality text generation.



The plots show steady decreases in both training and testing losses, indicating consistent learning over 10 epochs. Testing accuracy closely follows training accuracy, peaking at around epoch 8, suggesting good generalization with no significant overfitting. The small gap between training and testing metrics highlights the model's ability to generalize effectively to unseen data.