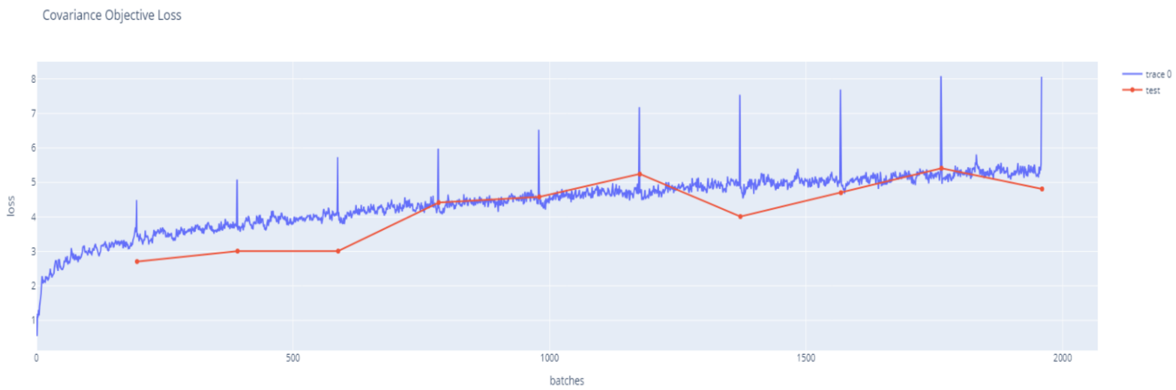
<u>67912 - Advanced Course in Machine Learning</u>

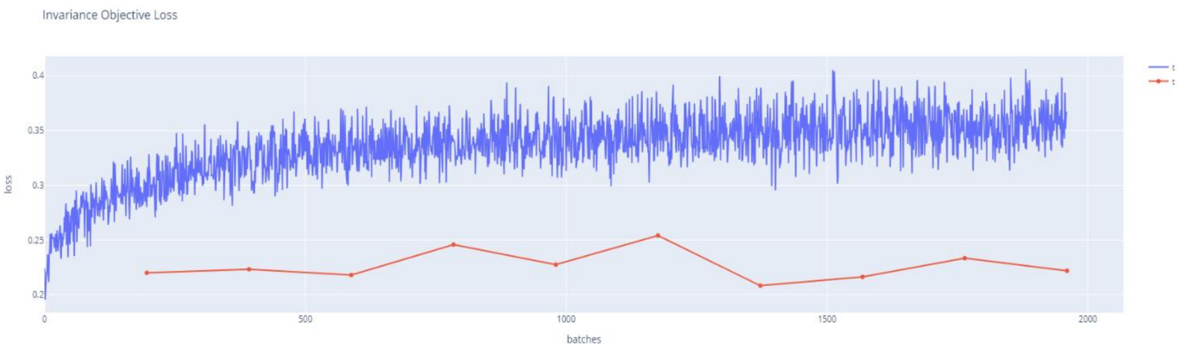
<u>Exercise 2 – Representation Learning</u>

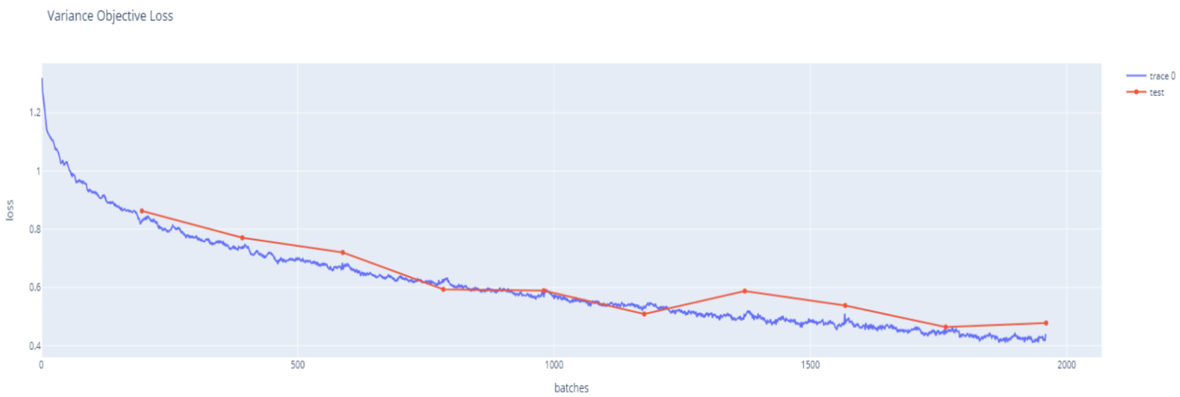
Gal Vardi, 208935031

3.2:

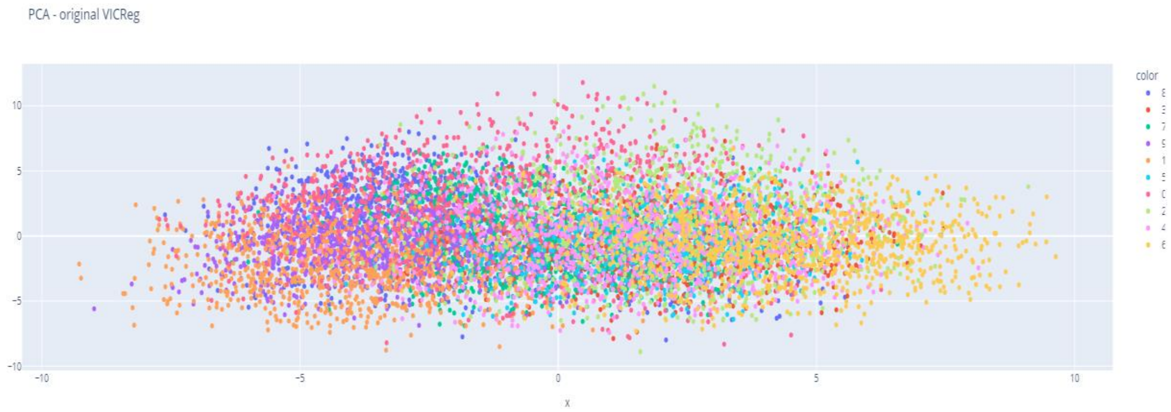
1.

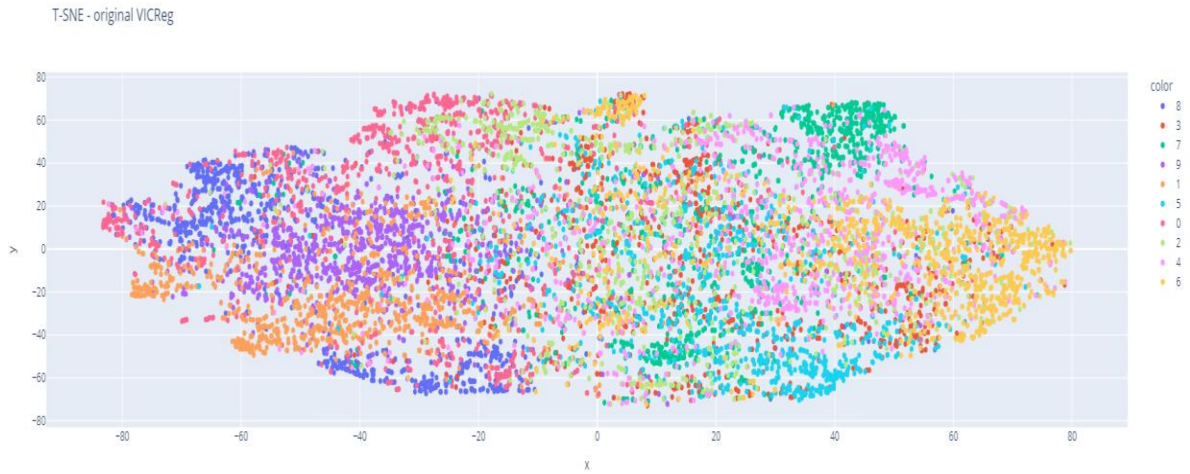






2.





In the T-SNE visualization the samples are more spread out . as well as It seems that the T-SNE representation was able to better differentiate between the classes which are less entangled than in pca, where most of the samples are in the dense mixed center. It appears that the model was able to capture the class information accurately as most of the samples of each class are clustered together. apart from classes 5,4,3,2 which seem especially entangled. Which might make sense as they are all embedding representing animals.

3.

```
original linear probing accuracy score -  0.6179
```

4.


```
no variance model linear probing accuracy score -  0.1255
```

As we can see the accuracy of the no variance model is much worse than the original, and only a bit better than random. As well as the fact that the pca visualization mapped every sample on one of two different lines. This may be explained by the fact that without the variance objective there is nothing to make sure each dimension is meaningful and so we get embeddings which are similar enough that pca maps them along two lines. Since the embeddings are similar (low variance) the task of classification based off of these embeddings is harder.

5.

The most prevalent problem in replacing VICReg's amortized encoder with LO, is that LO works by using an optimization process to find the best latent vector for each image. In doing so we lose the ability to train each vector (which now represent our embeddings) using each of our 3 loss objectives, and with it the advantages that each loss gives us. One possible solution would be to randomly initialize two vectors one for each view (augmentation) and using a generator optimize all three together, using reconstruction loss (to the two augmentations). As well as using contrastive loss where augments of the same image have a high probability of belonging to the same image, which will adjust their embeddings to be more similar and more dissimilar from vectors which correspond to a different image from their batch. The contrastive loss helps maintain VICReg's invariance objective, as well as increasing the variance between embeddings. Using these components and the reconstruction and contrastive losses could replace the original three loss objectives of VICReg and keep their benefits while using LO instead of an amortized encoder.

6.

```
no neighbors linear probing accuracy score -  0.4985
```

We can see that the accuracy of the no generated neighbors model is worse than the original model's accuracy . Since the original model is trained to be invariant to the differences between two augmentations of the same image, and in doing so 'hopefully' focus on the semantic features of the sample. I believe the difference can be explained by the fact that the invariance that the ablated model learns is not as subtle as augmentations on the same image since now the two views are of two distinct images and so some of the invariance that is learned is of features that may help characterize each class (the closest neighbor may be of a different class).

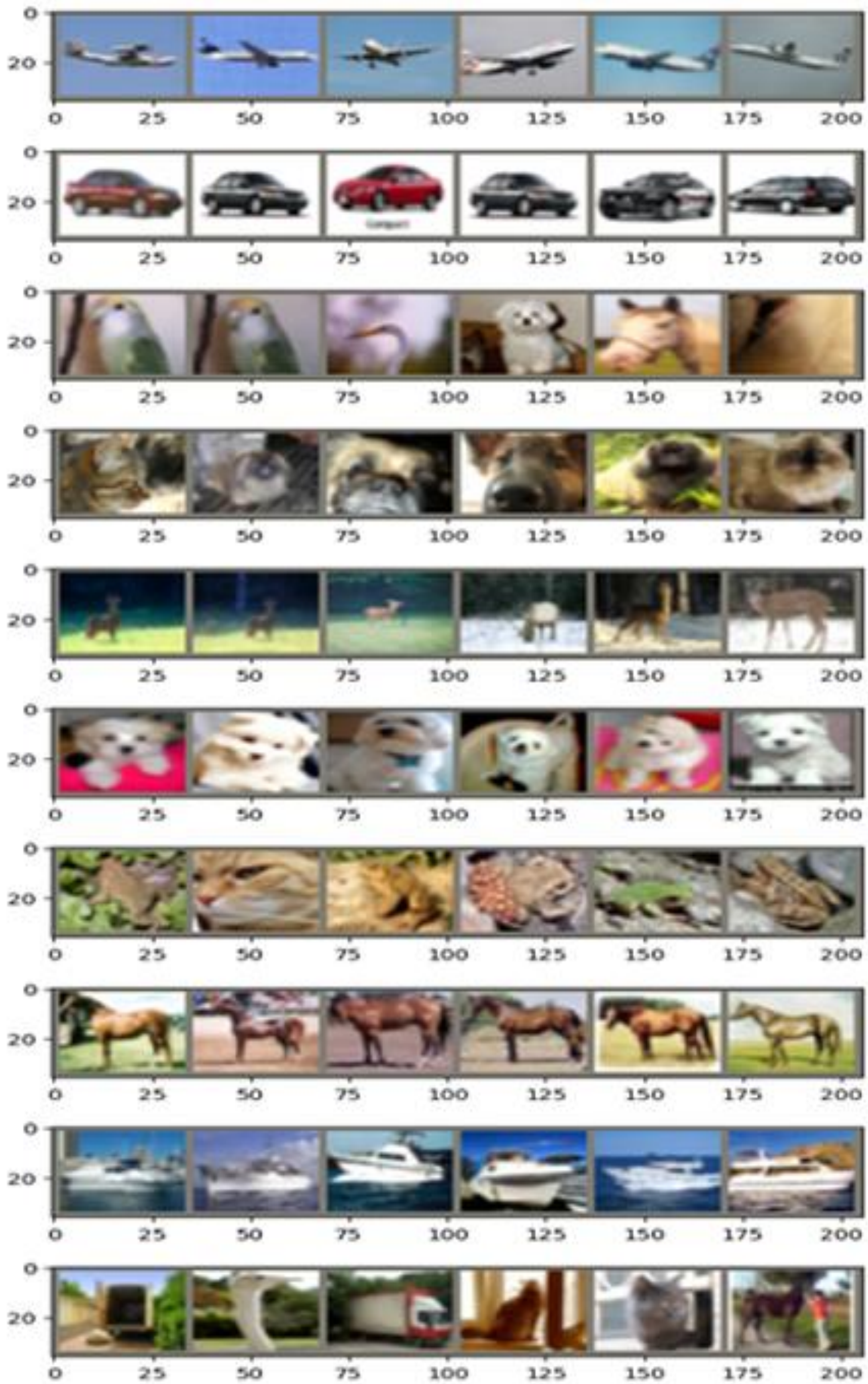
7.

We are able to see that in the T-SNE of the Laplacian eigenmaps embeddings all of the classes are mixed together with no real order of division between classes which is much worse than our original model in Q2. as well as the original model's accuracy being much higher 0.63 compared to 0.10. and so it appears that the Laplacian eigenmaps embeddings are not as distinctive from class to class as the original embeddings and so less effective for downstream applications. As we saw in the lecture the Laplacian eigenmaps algorithm minimizes the difference of embeddings based on distance L2. which is less ideal than generated neighbors as we saw in Q6. As well as the algorithm lacking a variance term which we learned of its importance in Q4. It may also be that some of the assumptions used for local embedding methods don't apply to the dataset such as lying on a 2d manifold.

8.

Original Model

Closest:



Farthest:

No Generated Neighbors Model:

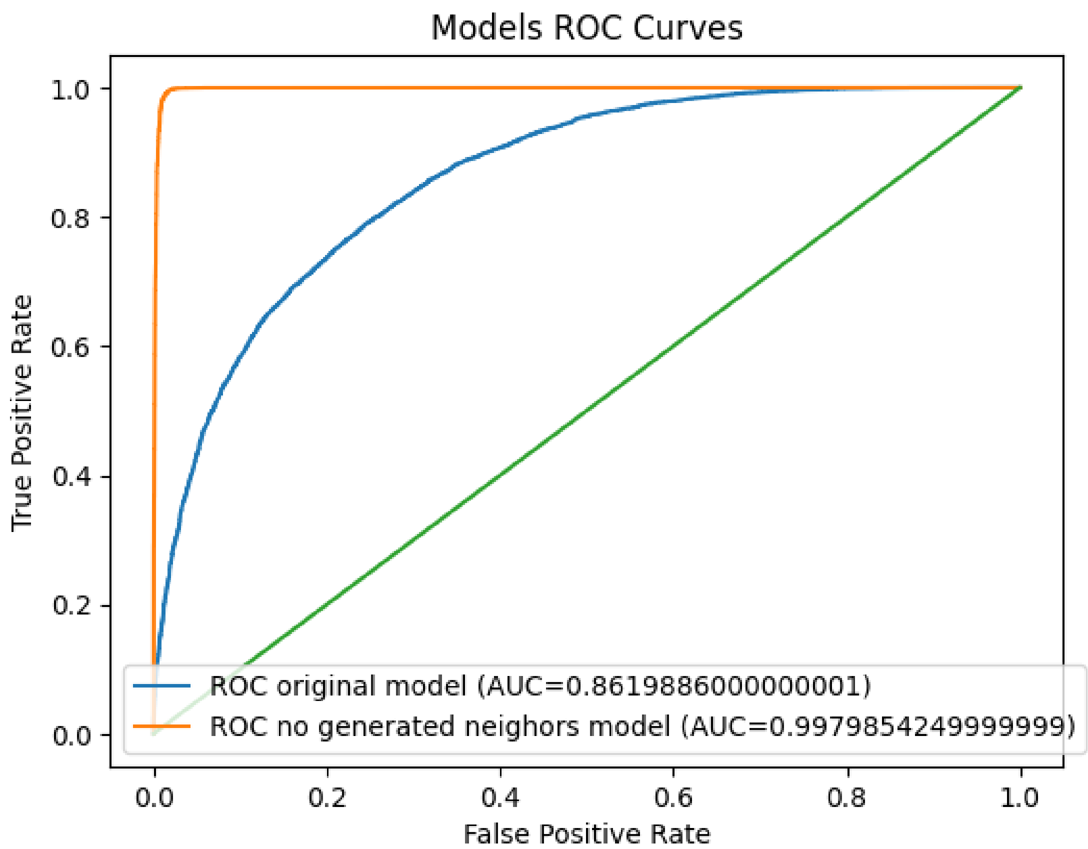
Closest:

Farthest:

It appears that the original model is noticeably better at maintaining close images together. although the no generated neighbors model's accuracy is good it has a harder time keeping close images together, as well as doing a slightly better job than the original at keeping distant images far apart. It seems that the no generated neighbors model is a bit more sensitive to color as a distinctive attribute as can be seen by its farthest images have a bit more different colors than the source. The original model doing better on the closest images could point to it doing a better job creating distinctive embeddings for classes which is corroborated by its higher linear probing score. as well as the ablated model using the same images for more classes than in the original further pointing to its inferior class distinction. It appears that in both models in the farthest images there are a few mutual images that are prevalent this may be because the neighbors for the training of the no generated neighbors model were picked based off of the embeddings of the original model.

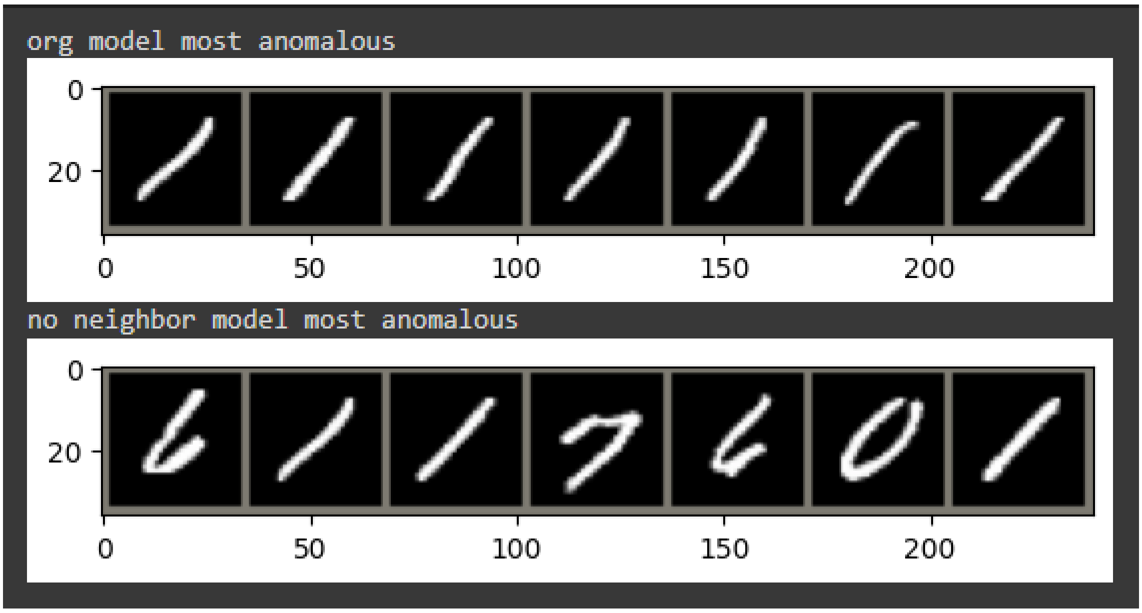
4.2.1 – Anomaly Detection:

2.



As we can see the no neighbors model is significantly better. I beleve this may be since in the ablated model the views were of two different real images from the training data the models embeddings may be more simular and in so more densely together in the embeding space which allowed the anoamlus mnist images to be more distant from them and recieve an a higher inverse score.

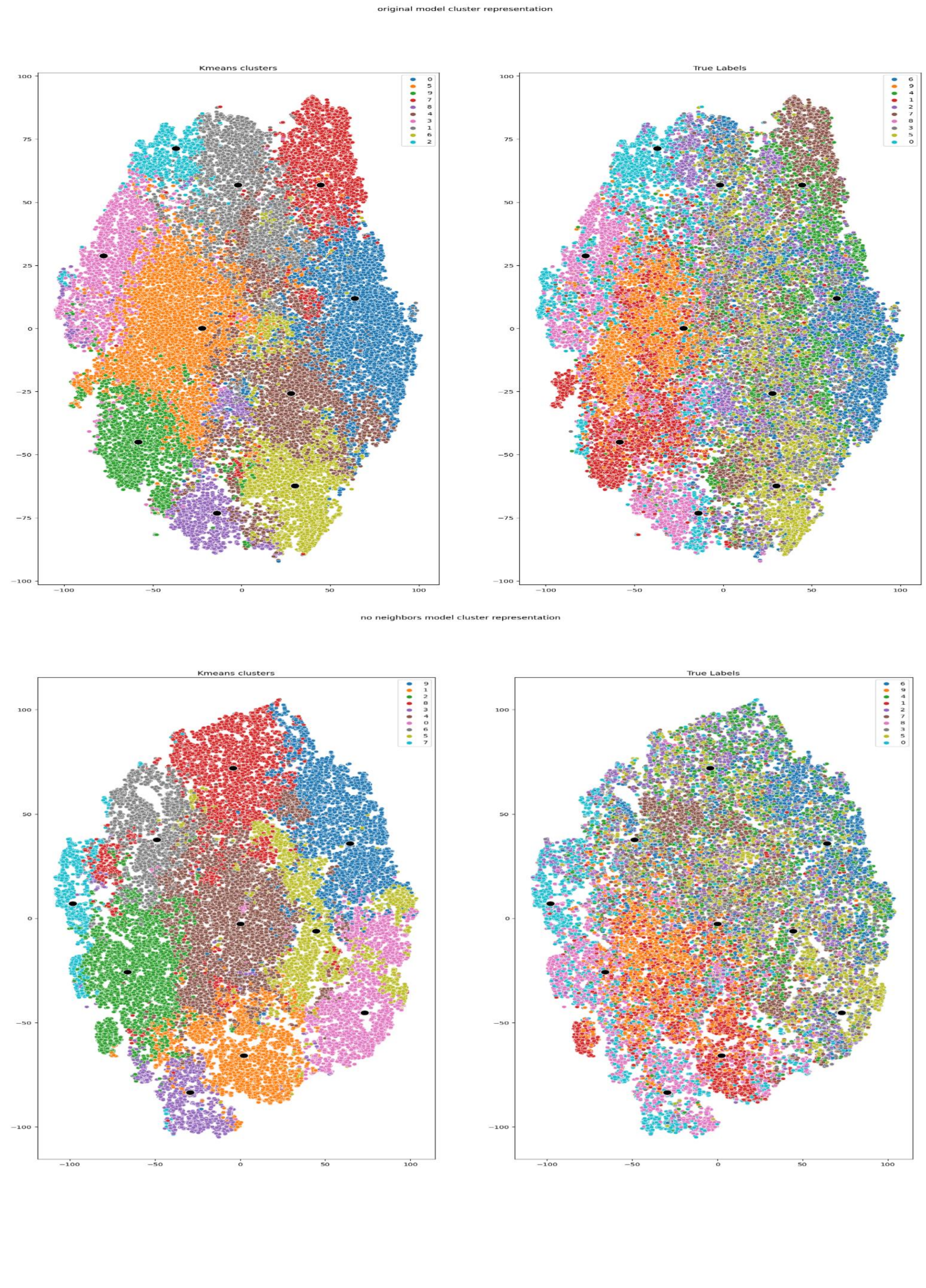
3.



Based on these results it is hard to tell which model is better at detecting anomalies given that they both choose images from the MNIST dataset. we a can see that the original models must anomalous images are mainly diagonal ones which could be interpreted as having the same diagonal attributes or colors as some planes as we saw in Q8. Which might further explain why it preformed worse than the second model, since it may have classified some planes as anomalous as well (for example). The second model chose more images with unnatural lines and curves which are less likely to be found in the original dataset. Further solidifying in my opinion that the no generated neighbors model is better at the given task.

<u>4.2.2 – Clustering:</u>

2.



It appears that the no generated model did a better job at finding clusters given that each class's centroid is more centered in regards to its samples and spaced in regards to other centroid. The original model seems better at separating between classes as they are less entangled. Which aligns with the fact that it accuracy in the linear probing was better at classifying different classes.

3.

Silhouette Scores:

```
Silhouette Scores:
orginal model 0.104025
no neighbors model 0.14165944
```

These results algin with the visual analysis that the no generated neighbors model does a better job at clustering. From looking at the silhouette score formula it is apparent that when b I larger than a the score is positive and an larger difference will converge to the score of 1. And so a higher score means that the points are overall more similar to others in their cluster than from other clusters.

<u>A.I Use:</u>

Throughout this project I used chatgpt and bard mainly for quick help with examples of uses of functions and tensor reshapes. As well as in the imshow function I used chatgpt to generate the unnormalize function to restore images to their original state for presentations sake after test_transform. I also partly used chatgpt to align the MNIST image size with the CIFAR10 images.