CIS 520, Machine Learning, Fall 2021
Homework 8
Due: Tuesday, November 9th, 11:59pm
Submit to Gradescope

Yixuan Meng, Zhouyang Fang

November 9, 2021

1 Performance Measures for Face Detection in Images

1. Based on the above numbers, calculate the TPR (recall), TNR, and precision of each group's system as tested. Also calculate the resulting geometric mean (GM) and F_1 measures for each system. If you were to select a system based on the GM measure, which system would you choose? Would your choice change if you were to select a system based on the F_1 measure? (Note, the geometric mean (GM) is defined as $\sqrt{TPR \times TNR}$.)

   Answer:

   (a) Group A:
   TPR(recall): TP/(TP+FN) = 2457/2700 = 0.91
   TNR(specificity): TN/(TN+FP) = 21357/(21357+243) = 0.98875
   precision: TP/(TP+FP) = 2457/(2457+243) = 0.91
   GM: $\sqrt{TPR * TNR} = 0.9486$
   F1: $\frac{2P*R}{P+R} = 0.91$

   (b) Group B:
   TPR(recall): TP/(TP+FN) = 276/300 = 0.92
   TNR(specificity): TN/(TN+FP) = 4428/(4428+72) = 0.984
   precision: TP/(TP+FP) = 276/(276+72) = 0.7931
   GM: $\sqrt{TPR * TNR} = 0.9515$
   F1: $\frac{2P*R}{P+R} = 0.8518$

   (c) Based on the GM measure, I would choose Group B, my choice will change to A if based on the F1 measure.

2. Which performance measure would be more suitable for this problem – the GM measure or the F_1 measure? Why?

   Answer: the F_1 measure. It balances the precision and the recall score. The group B method sets an unreasonable threshold and classifies too many regions as faces, the GM measure neglected this problem.

3. Another way to determine which method to choose would be to look at the ROC curve. Because you are given instances of different algorithms, not the algorithm itself, each method corresponds to a *point* on the TPR vs. FPR graph, not a curve.

   What is the Euclidean distance from each instance to the 0-error (perfect classification) point $(0, 1)$? Based on this metric, which method would you choose?

Answer: For group A, the TPR is 0.91 and the FPR (FP/FP+TN) is 0.01125. For group B, the TPR is 0.92 and the FPR is 0.016. The 0-error point is when the TPR is 1 and the FPR is 0.

The Euclidean distance from the 0-error point is 0.0907 for group A, and is 0.0816 for group B. Based on this metric, I will choose B.

4. Now assume that there is a third group C which trains their classifier using a newly discovered learning algorithm. Some of the resulting statistical measures are described below:

Research group	TPR	TNR	FPR	FNR
C	0.95	0.990	0.01	0.05

(a) Suppose you worked for a social media platform, where you have a large number of pictures, and want to avoid wrongly tagging some objects as faces. Would you prefer the algorithm created by group C over the algorithms created by groups A and B?

Answer: I will prefer the algorithm created by group C because the FPR is lower than the one of A or B.

(b) Now suppose you worked for law enforcement, where every face detected will be checked against a criminal database.You'd like maximize specificity TPR in this case to maximize the probability of finding the criminal. Would you prefer the algorithm created by group C over the algorithms created by groups A and B?

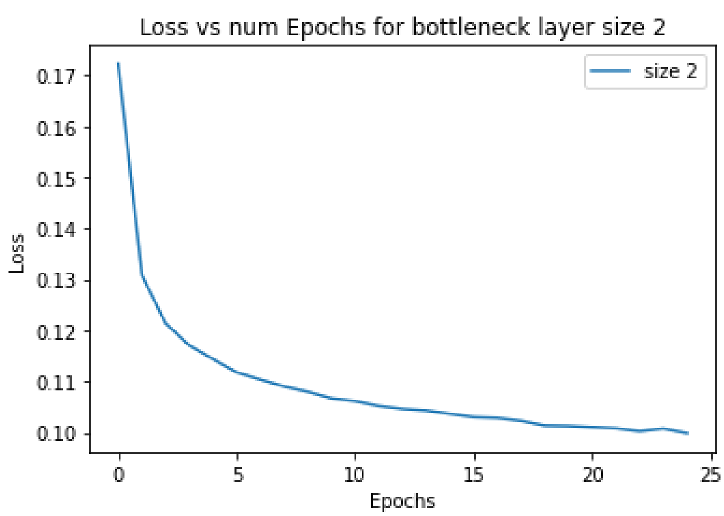
Answer: Yes, The TPR of C's algorithm is the highest.

2 Autoencoder

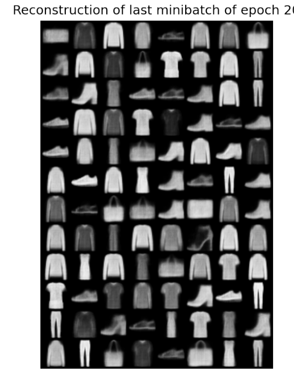
2.1 Part 1: Constructing the Autoencoder

1. Please report the reconstructed images from Epoch 0, Epoch 10 and Epoch 20. These reconstruc-tions are the output of your autoencoder for the last minibatch of the epoch. Record these outputs for0, 10, 20 Epochs respectively. Also report the training curve of the autoencoder (mean epoch loss vsepochs).

The training curve is given below:



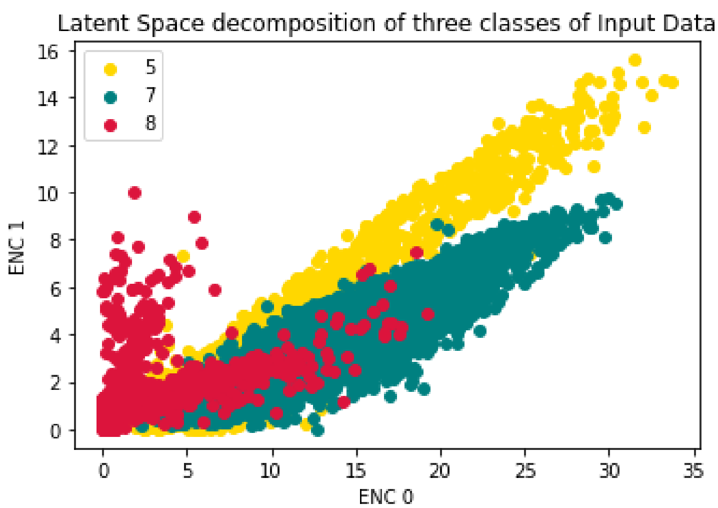
The reconstructions of the last minibatch are given below:



2.2 Part 2: Latent Space Decomposition

1. You will now plot the latent space of this autoencoder, which is essentially the encoding that is obtained for any given input. In this problem, the latent space is the output of the ReLU 3, after the Fully connected layer 3. Use your learned model from the previous section, and plot the images with the class label '5', '7', and '8' in the 2D latent space.

   The plot is given below



2. Based on the previous image, demonstrate both the difference and similarity among the patterns of label '5', '7' and '8' in the latent space decomposition plot. How are they different and/or similar? And Why? Recall the dataset label are: T-shirt/top 0, Trouser 1, Pullover 2, Dress 3, Coat 4, Sandal 5, Shirt 6, Sneaker 7, Bag 8, and Ankle boot 9.

   Answer: The patterns of label '5' and '7' in the latent space are similar in magnitude and direction, the pattern of '8', however, has a different shape. This makes sense because sandal and sneaker image are more similar, thus should result in more similar patterns in the latent space. Meanwhile, images for bags are dissimilar from the shoes, therefore the pattern of label '8' in the latent space looks different.

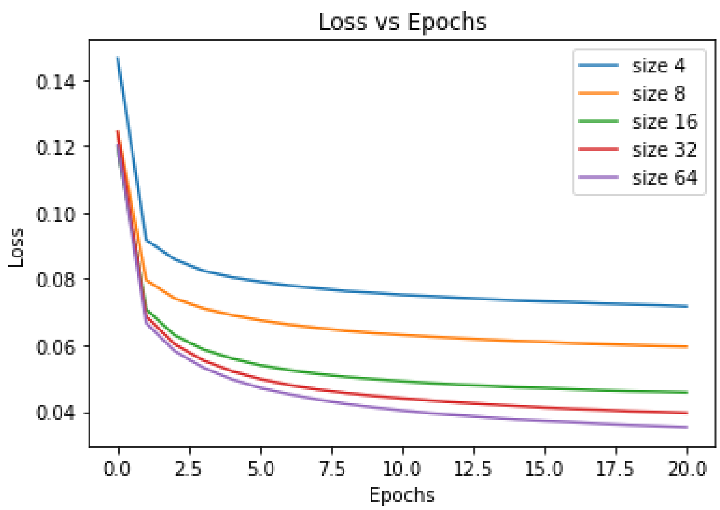
3. From the latent space plot, explain what the encoding has done to the inputs. How is this effect related to what PCA does? Why is this useful?

   Answer: To do the encoding, the encoder maps the data to a lower dimension space by extracting the vectors that can best represent the data. PCA also reduce the dimension of the feature space, though PCA is a linear transformation and autoencoders can do a non-linear one. They are useful because it compress the data without losing too much information.

2.3 Part 3: Reconstruction Error vs Bottleneck Layer

1. In this section, you will extend your code in Part 1 to a variable size bottleneck layer. Use your existing code in Part 1 and run the same for the bottleneck layer sizes $[4, 8, 16, 32, 64]$. Record the training curves and the final reconstruction errors for the input images belonging to the class label 5 for each of the sizes. Report the combined training curves (mean epoch loss vs epochs) for all the configurations. Also report the reconstructed images for Epoch 20 for each of the configuration.

   The training curves of all the configurations is shown below



   The reconstructions for sizes 4,8,16,32,64 for epoch 20 of the last minibatch are given below:

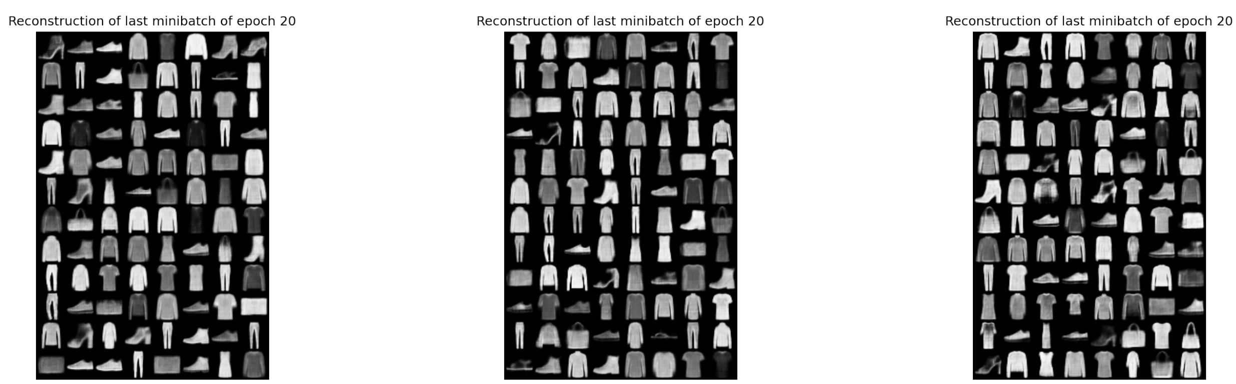
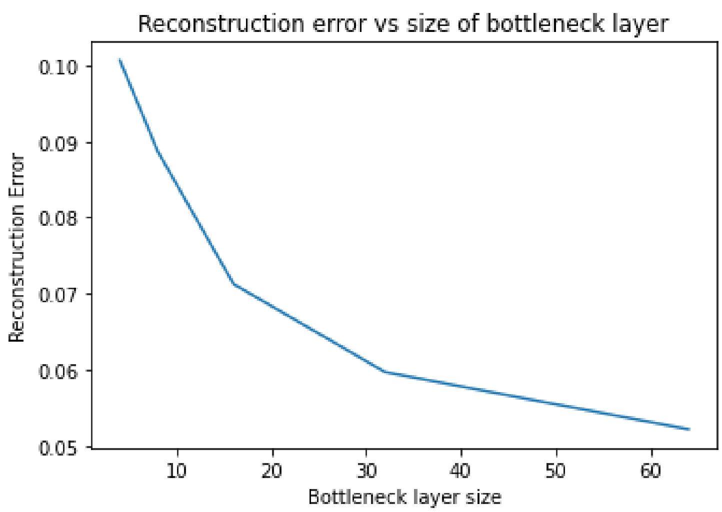


Figure 1: Size 4, 8, and 16

Figure 2: Size 32 and 64

2. Plot the mean reconstruction error of the 5 different trained models when the input images belong to the class label 5, with respect to the size of the bottleneck layer. What are you observing? How does what you see relate to PCA? What does this tell you about how you can potentially work with a high-dimensional data-space?

   The reconstruction error for images of class 5 as the bottleneck layer size increases is shown below:



   Answer: As the size of the bottleneck layer increases, the reconstruct error decreases; the decrease rate is higher when the size of the bottleneck layer is small. In PCA, the more principle component we use, the features in the latent space are more similar to the original ones. For a high-dimensional space, we can reduce it to a lower dimension without losing too much information by choosing a proper number of dimensions.