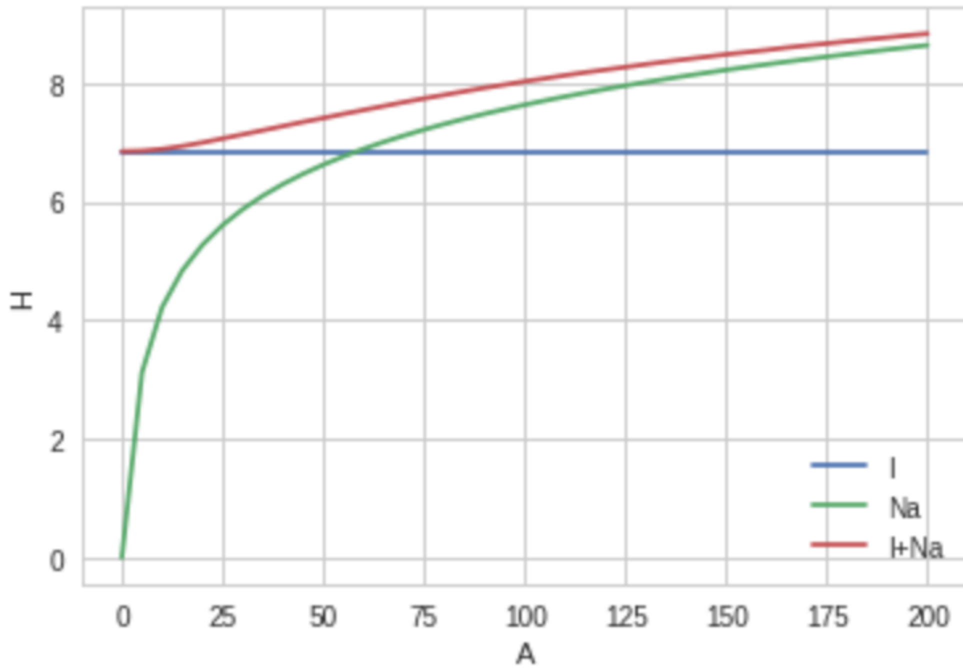
Grant Zhao
260563769

1.1.1

The entropy of I is 6.847, as shown in the juypter notebook.

1.1.2

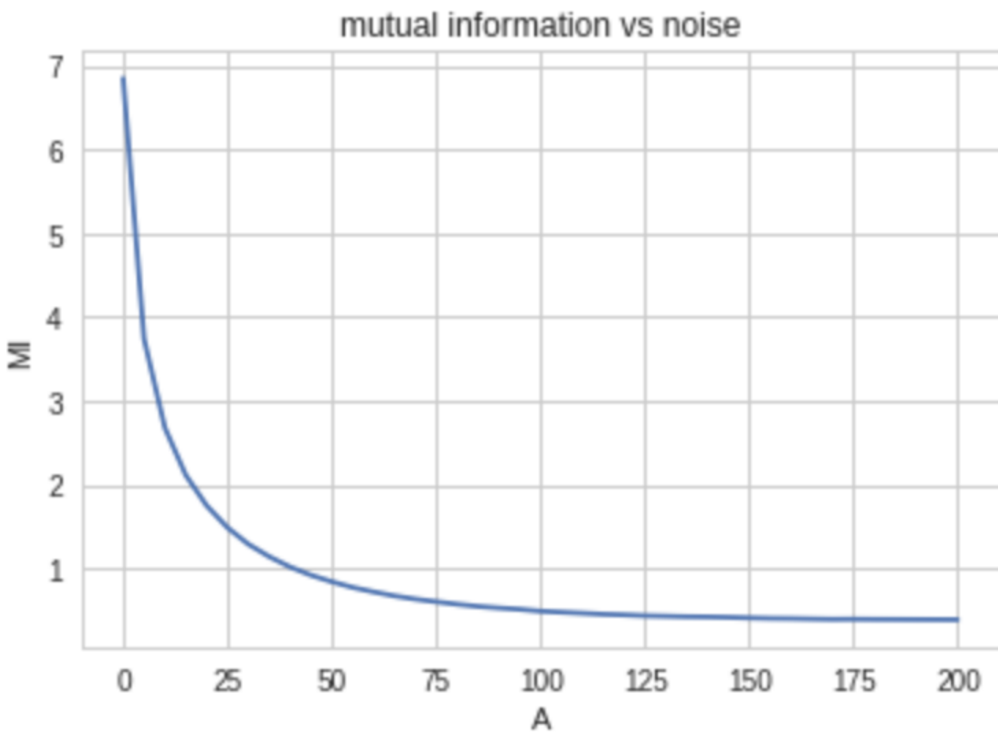
The plot of the entropy of I, NA and I + NA for multiple values of A is shown below. As we know, the entropy is related to the amount of uncertainty in the information source, it can serve a measure of disorder, as the level of disorder increases so does the entropy. In the plot below, as the intensity range of the noise rises, both the entropy of the image plus noise and the entropy of noise increases.



1.2.1

As shown in the diagram below, as the noise increase the mutual information between the image and the noisy image decreases. Mutual information measures the amount of information obtained about one image through observing to the other image, our plot make sense because as the range of the noise intensity gets larger, the information of the original image will be more covered.

Grant Zhao
260563769



1.2.2

In the notebook, it is verified that

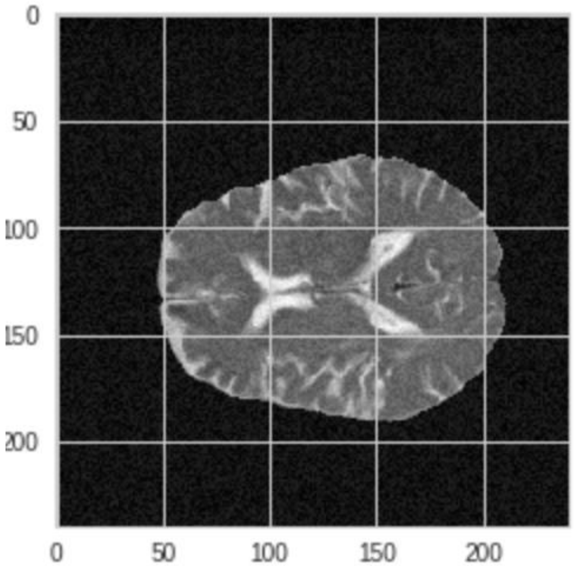
$$H(I; I + \mathcal{N}_{20}) = H(I) + H(I + \mathcal{N}_{20}) - MI(I; I + \mathcal{N}_{20})$$

1.2.3

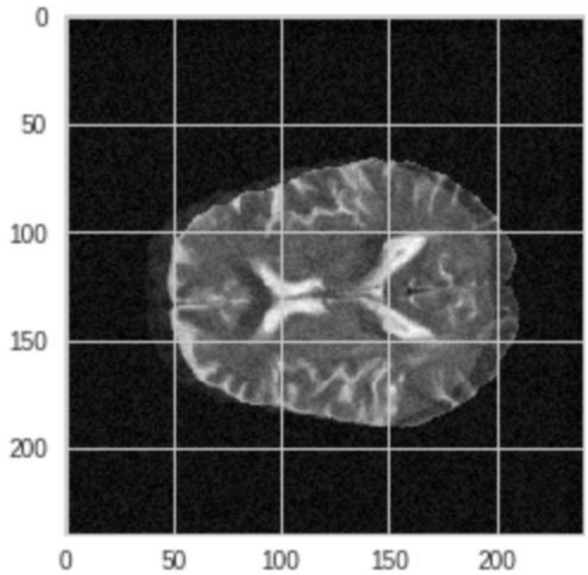
The forward KL divergence of the image and the noise is 1.1548, the backward KL divergence of the image and the noise is 2.16447. the forward KL divergence of the image plus noise is 0.224196, and the backward KL divergence of the image plus noise is 0.290273. As we know, the KL divergence measures the information loss when we use a approximation to model a distribution. As expected the KL divergence for the first group of images (image and pure noise) is much larger than the second group (image and image plus noise). And the forward and backward KL divergence for each group are similar, because change roles between two images will not cause extra information loss.

1.3.1

For the first group of images, there is a displacement of 3 to the x direction 10 to the y direction between the image I1_1 and I1_2, according to Mutual Information. And -8 to the x sirection 9 to the y direction according to MSE. The visualizations of the shifting is shown below.

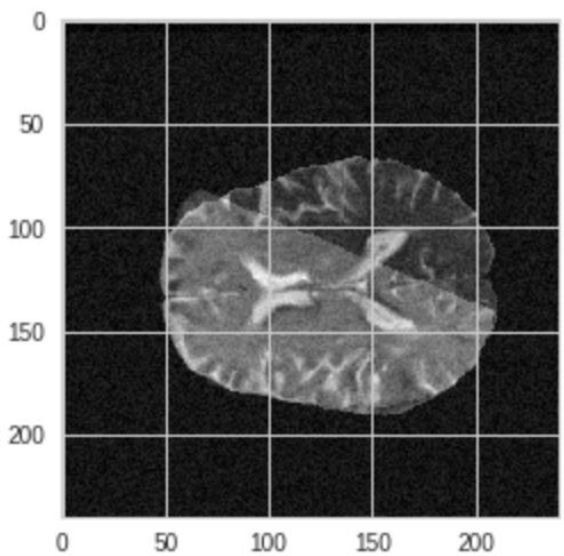


Mutual information

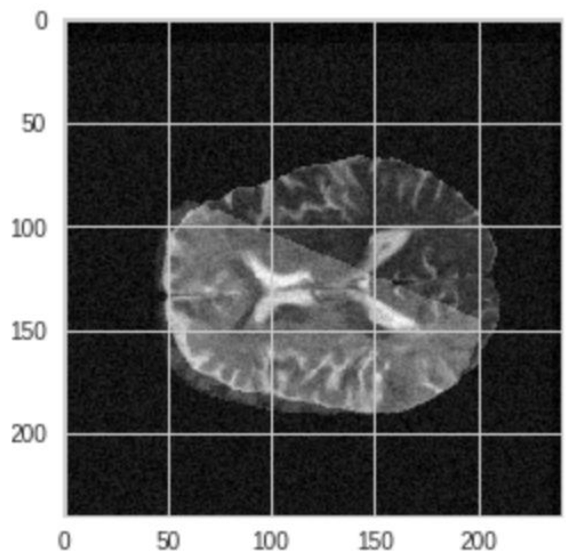


MSE

For the second group of images, there is a displacement of 3 to the x direction 10 to the y direction between the image I2_1 and I2_2, according to Mutual Information. And -8 to the x sirection 9 to the y direction according to MSE. The visualizations of the shifting is shown below.



Mutual information



MSE

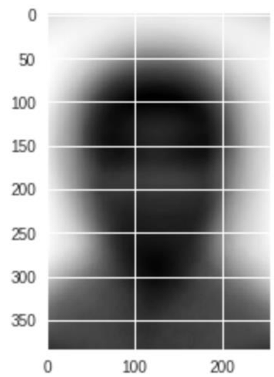
1.3.2

The MSE method is trying to match the pixel intensities of one pixel on one image to another pixel at the exact location of another pixel. Thus it is important for the object in the two images to have the same shape and pixel intensity, it will not perform well when an object in one image has one color but the other object has another. MI method will not perform well when there is a lot of back ground noise, background noise is extra information when calculating the entropy. Thus, when there is a clean background like pure color, mutual information should be used. And when the background contains noise, but the objects are the same, MSE should be used.

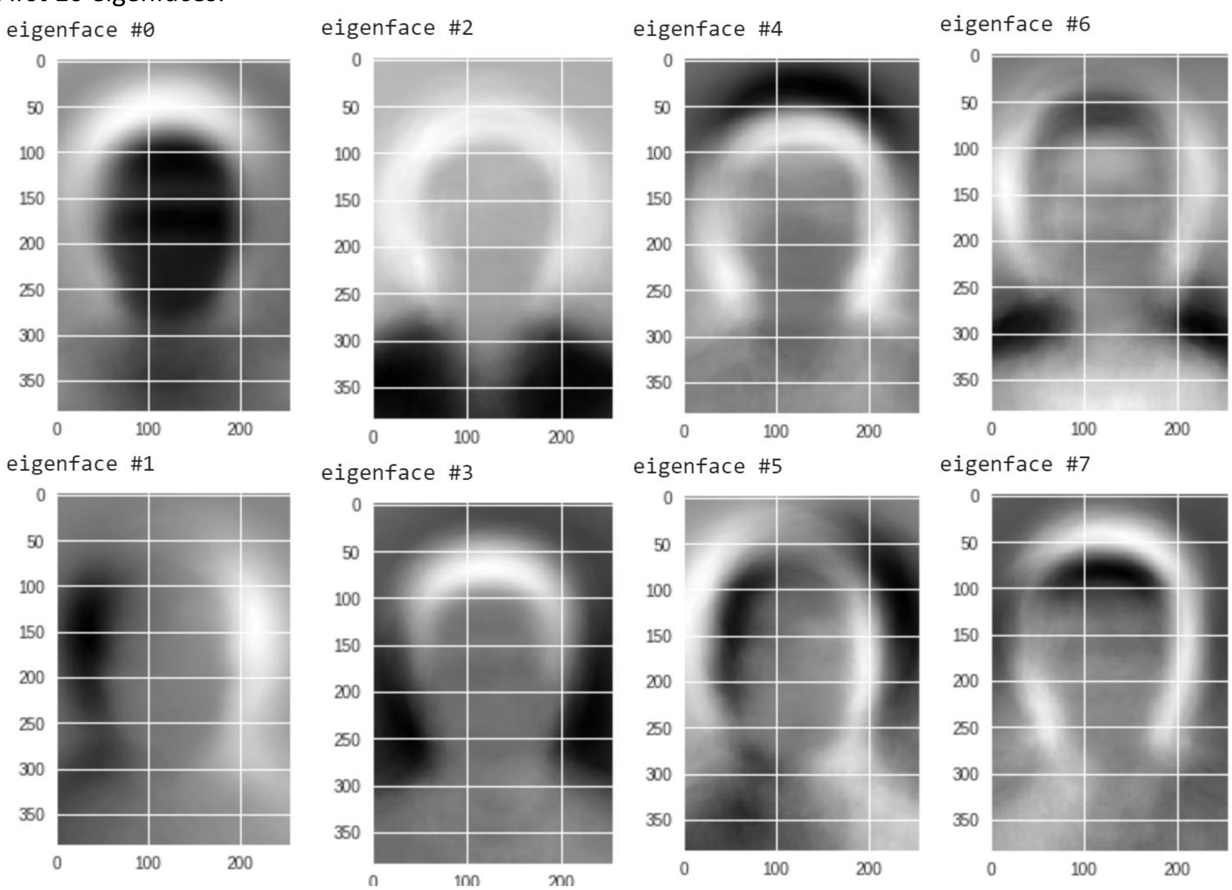
Grant Zhao
260563769

2.1 PCA

Mean face:

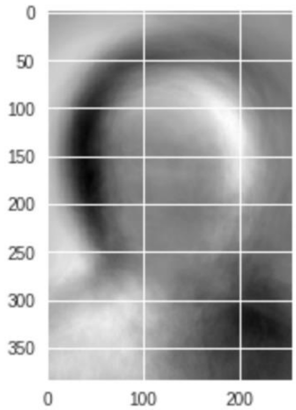


First 10 eigenfaces:

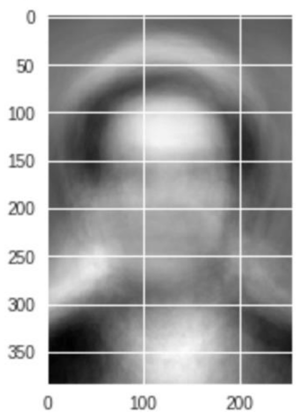


Grant Zhao
260563769

eigenface #8



eigenface #9



2.1.2

I used the fraction of total variance approach, the fraction of the total variance represented by the first N eigenfaces is given by the fraction of the sum of their corresponding eigenvalues out of the sum of all of the eigenvalues. According to my program the first 17 eigenfaces represent more than 75% of the variance and the first 48 eigenfaces represented more than 85% of the total variance. I chose 20 as my eigenface number, which should approximately represent 80% of the total varance.

2.2.1

We denote the posterior probability of a class C_i given observation y (eigenfaces feature) as $\mathrm{P}(\mathrm{C}_i|\mathrm{y})$, then by Bayes Theorem we have:

$$\mathrm{P}(\mathrm{C}_i|\mathrm{y}) = \frac{P(y|\mathrm{C}_i) * P(\mathrm{C}_i)}{P(y)}$$

Where $P(\mathrm{C}_i|y)$ is the likelihood function which is the probability of the observation y given the class C_i. The observations are eigenfaces and we assume the eigenface features to be conditionally independent. The likelihood function is equal to $P(y|\mathrm{C}_i) = \prod_j^n P(\mathrm{y}_j|\mathrm{C}_i)$ where n is the number of eigenfaces. We know that P(y) is a constant across all class, the face recognition problem can be solved by finding the class that will maximize the posterior function $\mathrm{P}(\mathrm{C}_i|\mathrm{y})$.

2.2.2

The eigen representation is built in the juypter notebook.

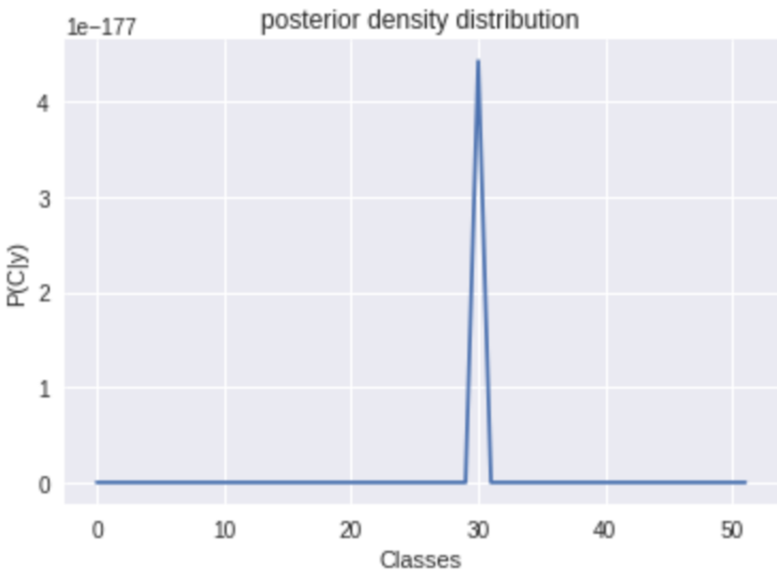
Grant Zhao
260563769

2.2.3

The density functions are built in the juypter notebook.

2.2.4

The diagram below is a visualization of one of the test image's posterior density distribution, apparently class 30 maximize the posterior, and class 30 corresponds to identity 760 which matches with the true identity.




```
the perdicted label is 760
the actrual label is 760
```

2.2.5

The accuracy of the two methods are similar, while the nearest neighbour method appears to give better result. The MAP method has an accuracy of 50.25% and the nearest neighbor method has an accuracy of 53%.