Week 5

Progress during July 27th-Aug 2nd

Hao SUN

August 2, 2017

Contents

1. Unsupervised learning in star-galaxy classification/segmentation
   - Tried to find the correspondance of the hidden variables
   - Used manifold learning and priori knowledge get an unsupervised classification accuracy of 79.0%
   - Segmentation and classification
   - Redshift prediction

Progress in this week

1. Tried to find the correspondance of the hidden variables
2. Used manifold learning and priori knowledge get an unsupervised classification accuracy of 79.0%
3. Performed Hypercolumns with manifold learning result to perform both segmentation and classification
4. Redshift prediction, validatioin MSE $= 0.0025$

Tried to find the correspondance of the hidden variables

Used naive method to calculate the size and magnitude of objects. And compare them to the hidden variables VAE found.

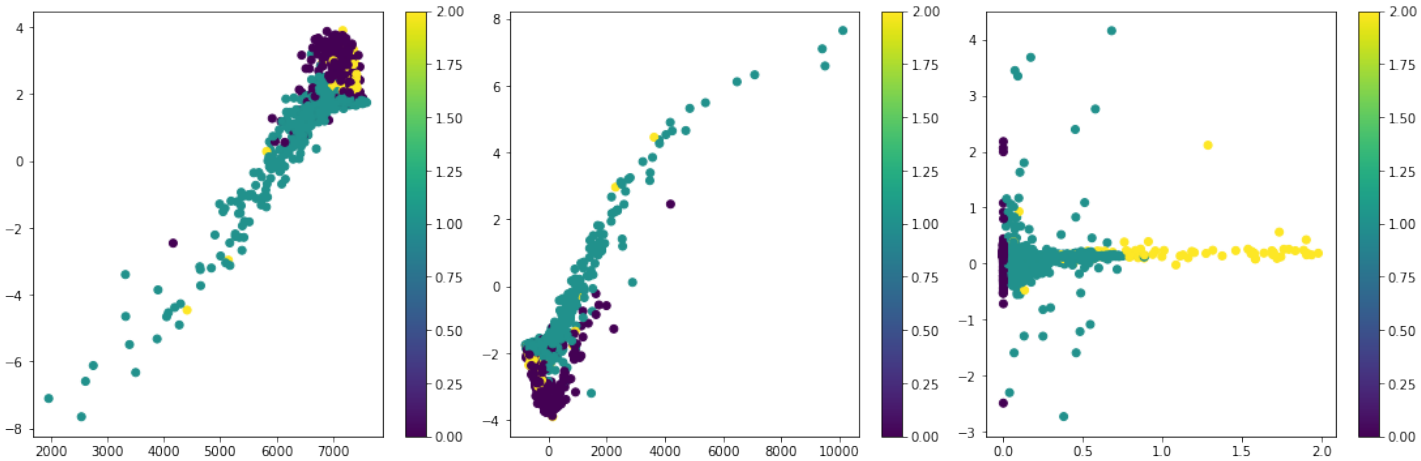


Figure: hidden variable vs. size(L), magnitude(M) and redshift(R)

Used manifold learning and priori knowledge get an unsupervised classification accuracy of 79.0%

In unsupervised learning, for example, VAE, the neural net learned about useful features it may found to minimize the lossfunction (To make the generated images look like the original images). Without labels, the classification result maybe not what we want but what VAE found.

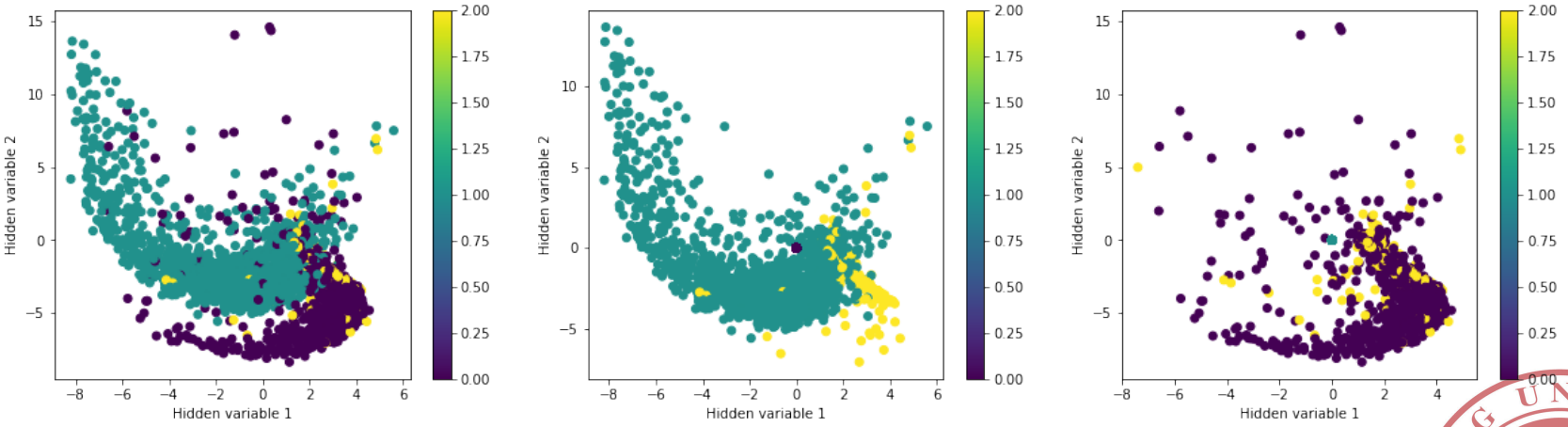


Figure: hidden variables, VAE gets no motivation for improvement

Kmeans, Density based and Manifold learning clustering

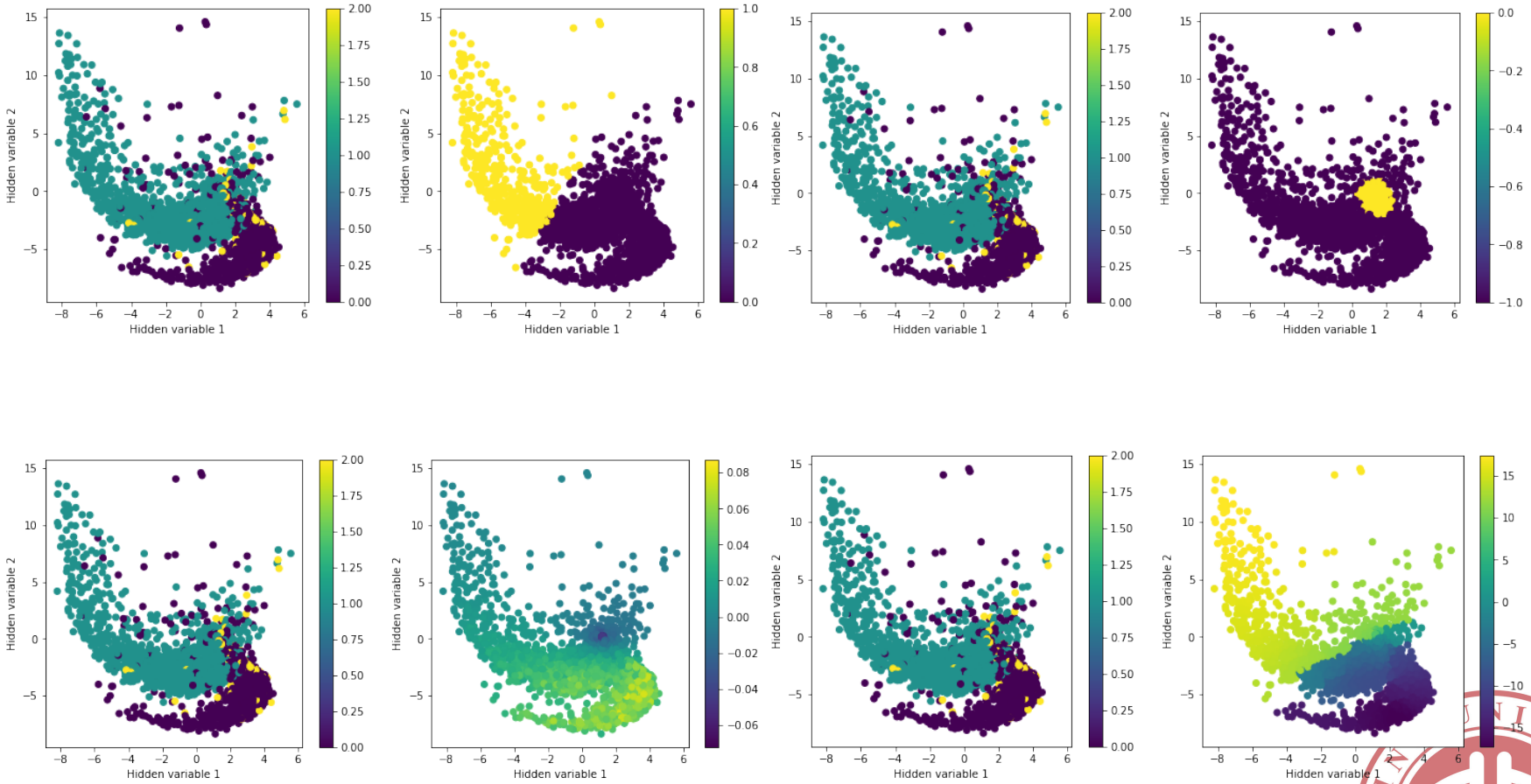


Figure: Kmeans(L;U), DBSCAN(R;U), SpectralEmbedding(L;D) and t-NSE(R;D)

ISOMAP + priori knowledge

The propotion between galaxies and stars(+Quasars) are about 2:1. Use this priori knowledge to find the threshold in manifold clustering to perform classification.

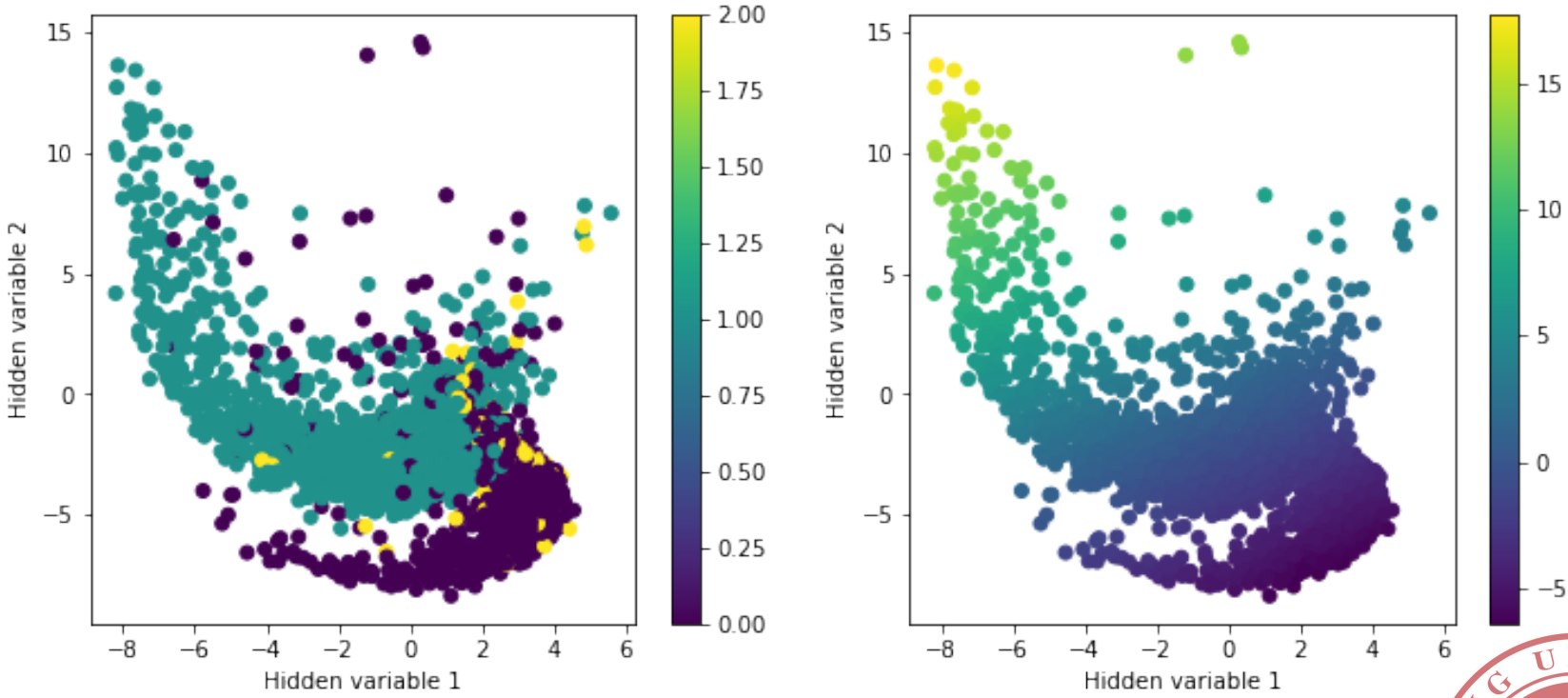
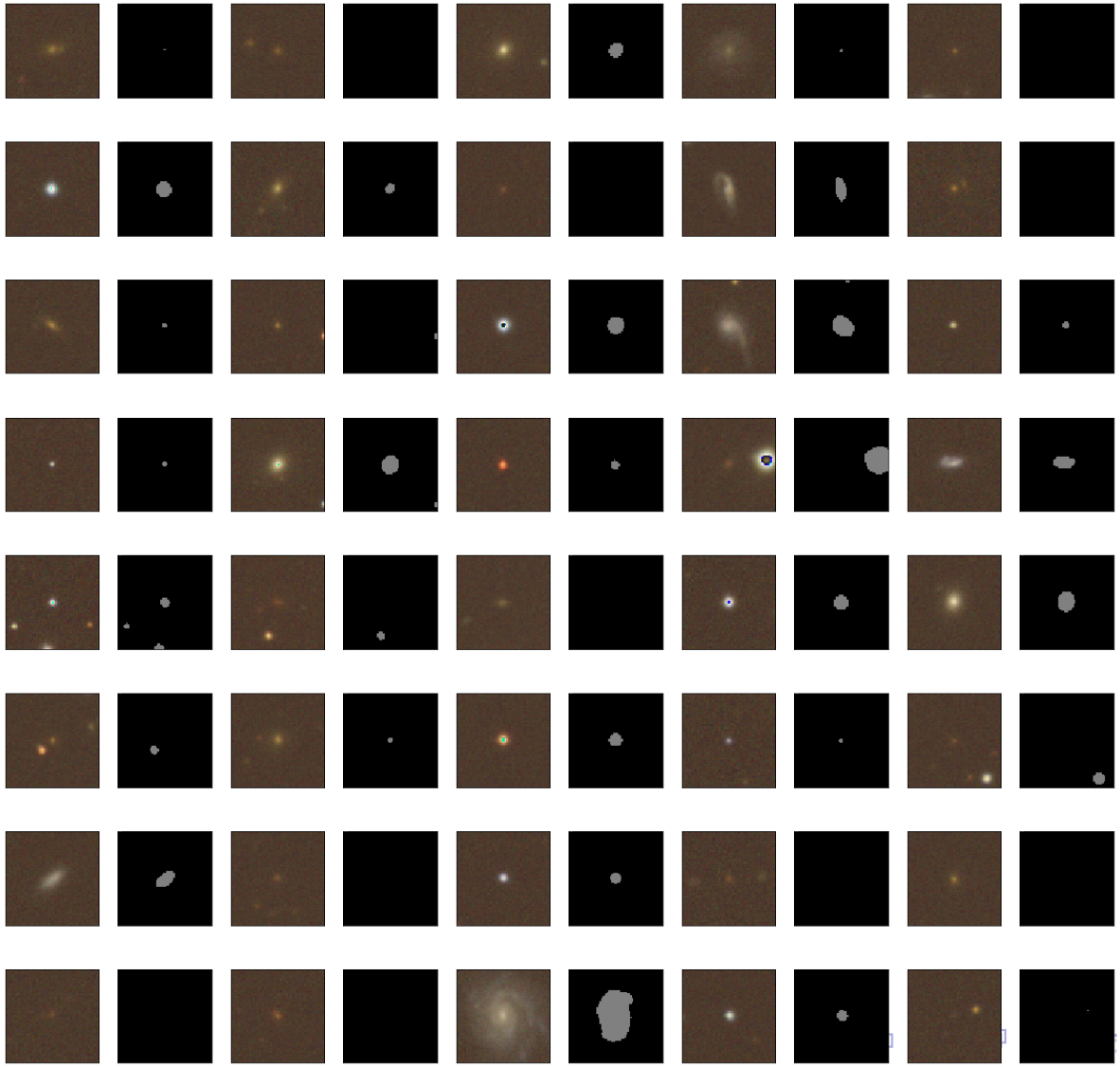


Figure: ISOMAP clustering

The classification accuracy is 79.0%(82%&66% for galaxies and stars)

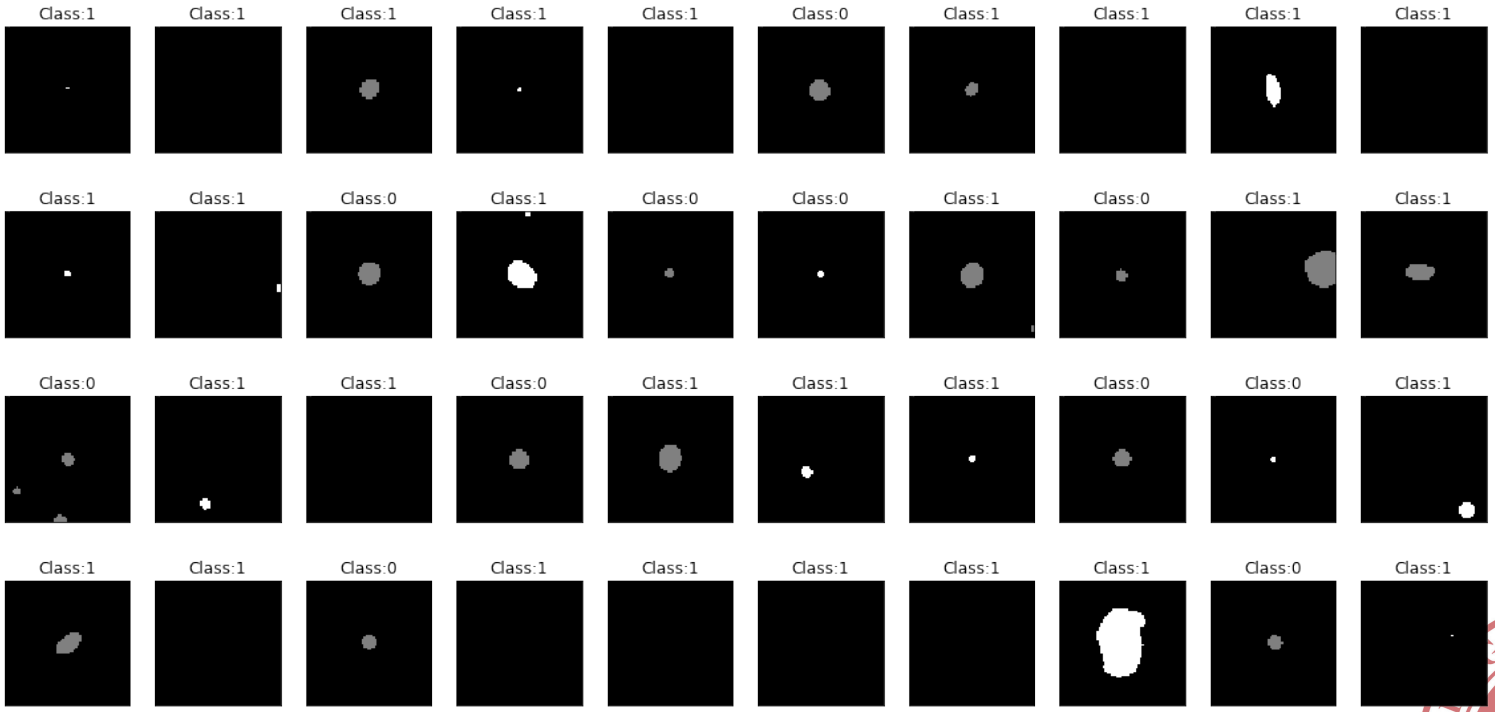
Segmentation

Use the first few layers with Hypercolumns, we can perform segmentation



Classification

Use the first few layers and the hidden layer in Hypercolumns, we can perform both segmentation and classification



Accuracy of those 40 images is 70%

Redshift prediction

Find a paper on this topic. arXiv: 1703.01979v1. The paper used DCMDN(Mixture Density)
I tried to use BN + ConvNet + LeakyReLU and got a validation MSE of 0.0025(*2.56)

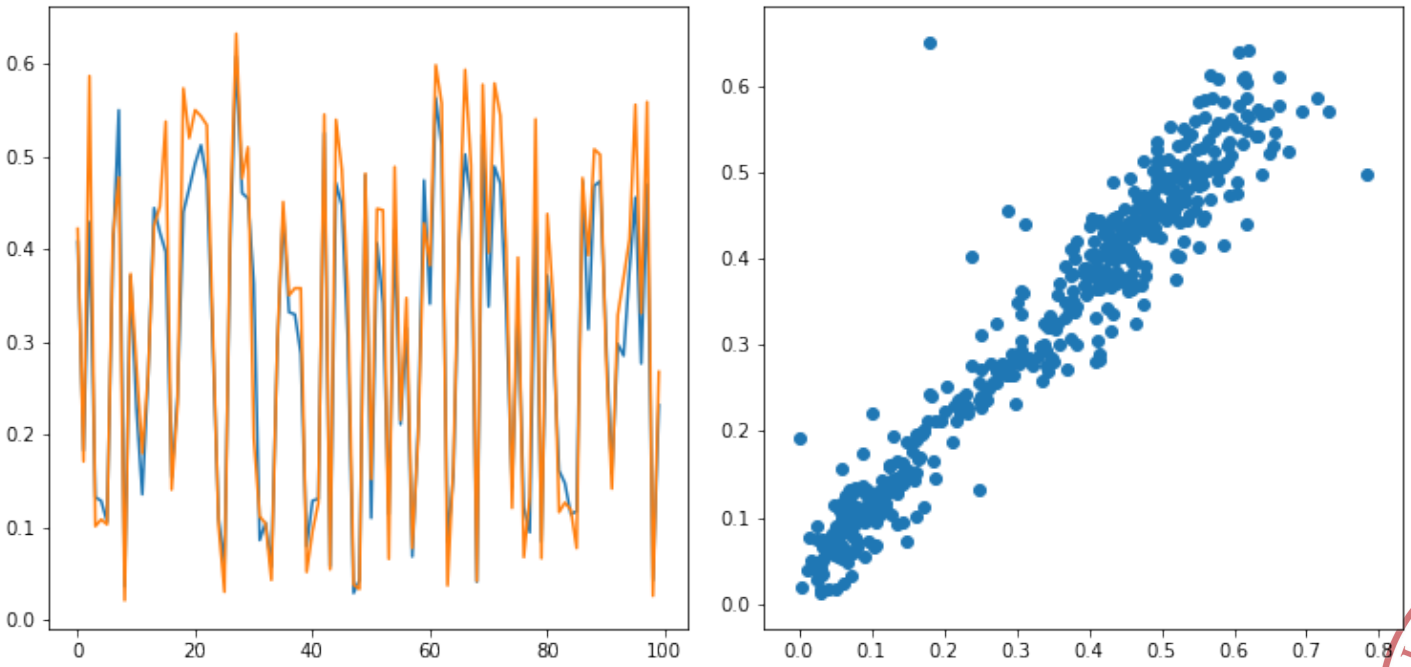


Figure: L: pred(blue) and real(orange) redshift on part of validation set, R: pred(Y-axis) and real redshift(X-axis)