Detection of Exoplanets from gaps in Protoplanetary Disks

CS460 - Midterm Report

Anshada P M

Varun M.

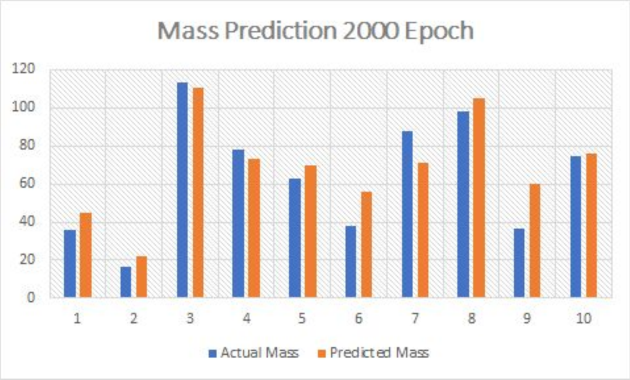
The Progress so far:

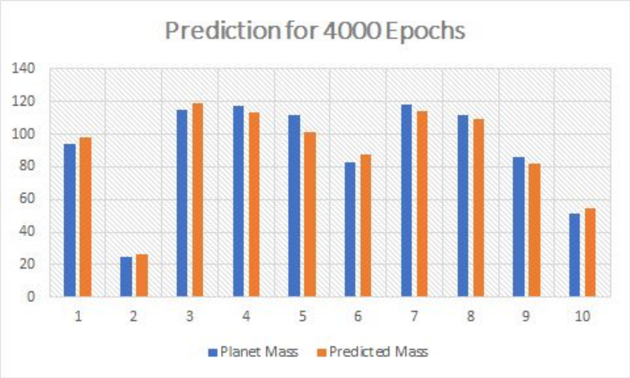
So here's the update:

- The project involves **prediction of exoplanets** in young stars by **detecting gaps** in protostellar disks. As mentioned before the origin of these disks can be due to many reasons. We have been trying to build this by mainly using **Tensorflow keras** since the algorithms available allow us a wide range of functions.
- The project first involves the designing of a **CNN for image recognition** from the image data which has been obtained from **ALMA.**
- We are working on two datasets simultaneously, on an already preprocessed dataset and on the feature engineering of the images by ALMA.
- Codes developed were tested on the preprocessed dataset to check for performance.
- We created a **Multilayer Perceptron** to predict Planet Mass and are trying to use Isolation forest to separate disks which we assume have gaps due to other reasons.

Planet Mass Prediction

- This prediction is done using **Keras Sequential API**. For this we used deep learning consisting of 3 deep layers consisting of 256, 128 and 1 nodes.
- For the model the activation **function relu** was applied and **L2 regularisation** was used.
- The optimizer used is the **RMSprop algorithm.**





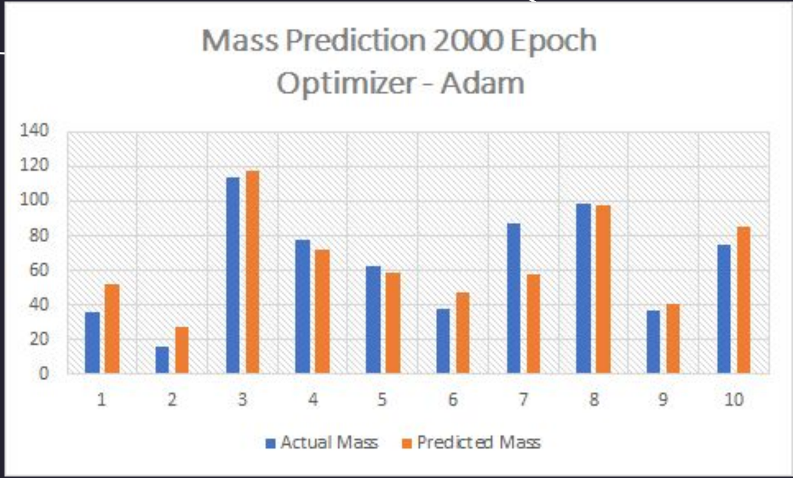
The Base Code for Planet Mass Prediction -1

```python
def build_model():
    model = keras.Sequential([
        layers.Dense(256,activation=tf.nn.relu,kernel_regularizer=regularizers.l2(0.0001),input_shape=[len(train_dataset.keys())]
        layers.Dense(128,activation=tf.nn.relu,kernel_regularizer=regularizers.l2(0.0001)),
        layers.Dense(1)
    ])
    optimizer = tf.keras.optimizers.RMSprop(0.0001)
    model.compile(loss='mean_squared_error',
                  optimizer=optimizer,
                  metrics=['mean_absolute_error', 'mean_squared_error'])
    return model
model = build_model()
model.summary()
```

	loss	mean_absolute_error	mean_squared_error	val_loss	val_mean_absolute_error	val_mean_squared_error	epoch
1995	79.261093	6.464504	79.163971	199.589096	10.070127	199.492050	1995
1996	79.910500	6.435901	79.813393	198.290207	10.044401	198.193039	1996
1997	79.292183	6.412332	79.195038	197.820358	10.045137	197.723160	1997
1998	79.701218	6.468977	79.604042	197.842590	10.047734	197.745361	1998
1999	79.280098	6.459575	79.182869	199.048889	10.063108	198.951706	1999

Using Optimizer - Adam

Upon using optimizer function **Adam.** For this optimizer we get very low error training error as indicated.



	loss	mean_absolute_error	mean_squared_error	val_loss	val_mean_absolute_error	val_mean_squared_error	epoch
1995	3.812508	1.483621	3.589180	231.846786	11.726738	231.623505	1995
1996	2.157659	1.070647	1.934400	233.237381	11.694411	233.014114	1996
1997	1.573927	0.860021	1.350678	227.963730	11.478926	227.740524	1997
1998	1.685055	0.890177	1.461808	233.529984	11.602855	233.306732	1998
1999	1.569372	0.838148	1.346102	229.171310	11.508163	228.948013	1999

Creation of Isolation Forest

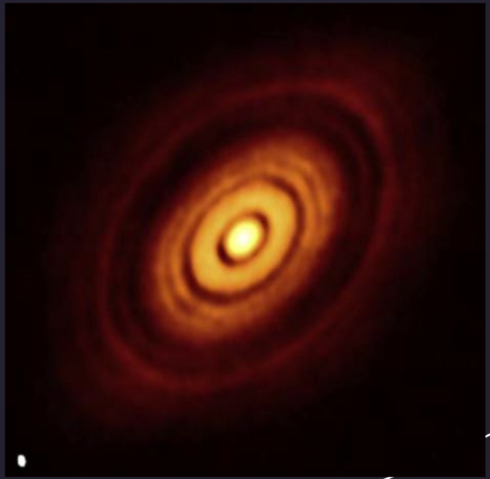
- Isolation forest algorithm has been used to attempt to classify whether the disks are host to planets or not.
- This has been implemented using sklearn library. Functions used include **make_pipeline , StandardScaler, GradientBoostingRegressor and IsolationForest.**
- Anomalies popped up as -1 and were compared for efficiency by looking at **Dust Gap as the parameter**. Initial efficiencies are low and we plan to improve it by tweaking the parameters accordingly. Imports parts of the code are depicted below:

```
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.ensemble import IsolationForest
model = make_pipeline(StandardScaler(),GradientBoostingRegressor(random_state = 0, n_estimators =700, max_depth = 6, min_samples
model.fit(y_train,X_train)
print(model.score(y_train, X_train))

model = IsolationForest(n_estimators = 700,max_samples = 'auto',contamination = float(0.3),max_features =1.0  )
model.fit(dataset[['Planet_Mass']])
```

Image Detection of Protostellar Disks

- For CNN which we are currently trying to use the Network type **RESNET50** whose structure we have studied. Possible use of **ALEXNET** if we can obtain better results.
- For optimizer in keras we are using **Adam** which uses a stochastic gradient descent method. We might try **Adadelta** if time permits.

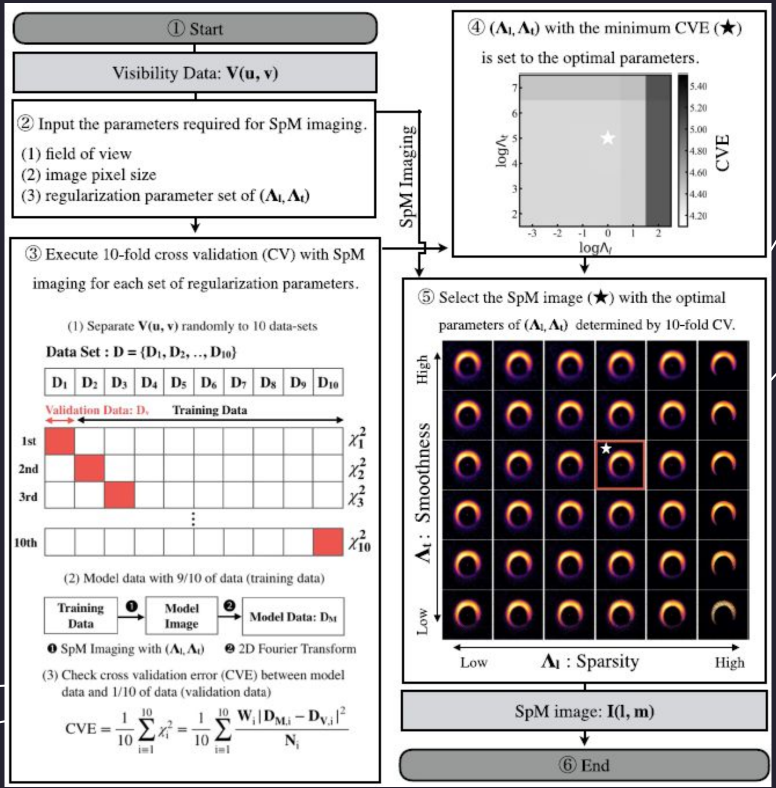


Contribution From Papers

- Atacama Large Millimeter Array (ALMA) has enabled us to observe Protoplanetary Disks (PPDs) with good resolution. But this resolution has been observed not to be adequate enough. Hence before running a CNN we need to use a **super-resolution imaging techniques** to be able to observe the disks.
- In one of the papers they utilised **sparse modelling (SpM)** to enhance the resolution of the image by **30%** compared to the conventional algorithm used. Since developing a new super resolution algorithm is out of the scope we will try to obtain similar algorithm to increase resolution so that we can run the CNN.
- Using SpM a resolution of **5 × 4 au** was obtained using which they observed disk and ring structures of the a **young triple system T Tau** from the ALMA archival data.

Sparse Modelling

- The SpM image is generated upon giving a user specified parameter.
- The user specified parameters are **field of view, image pixel size, regularization parameters.**
- Cross validation error is calculated and the parameters with the **minimum CVE** is set as the optimal regularization parameters.
- SpM image with the optimal regularisation parameters is selected.



$$CVE = \frac{1}{10}\sum_{i=1}^{10}\chi_i^2 = \frac{1}{10}\sum_{i=1}^{10}\frac{W_i\,|D_{M,i} - D_{V,i}|^2}{N_i}$$

References

https://iopscience.iop.org/article/10.3847/1538-4357/aba95d/pdf-A Machine Learning Model to Infer Planet Masses from Gaps Observed in Protoplanetary Disks

2107.09086v1 [astro-ph.EP] 19 Jul 2021 - Finding hidden planets from simulated images of protoplanetary disk gaps

https://arxiv.org/abs/2110.00974 - ALMA Super-resolution Imaging of T Tau: r = 12 au Gap in the Compact Dust Disk around T Tau N.

https://iopscience.iop.org/article/10.3847/1538-4357/ab899f/pdf -Super-resolution Imaging of the Protoplanetary Disk HD 142527 Using Sparse Modeling