AstroML

Machine Learning and Data Mining for Astronomy

Examples

Simple Gaussian Naive Bayes Classification

- Using multivariate normal distribution data
- Convolution Features data of SuperNova Remnants (SNR) from Chandra Supernova Remnant Catalog.

Artificial Neural Network Estimation

- Using SDSS Galaxies with spectral information

Overview - Gaussian Navie Bayes Classification

Reproduce

First we reproduce example present on AstroML homepage.

Apply

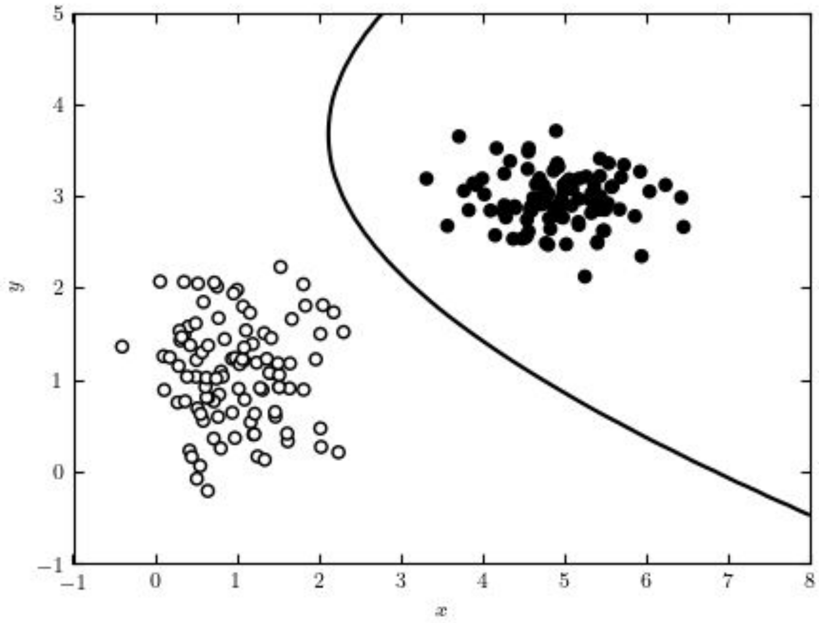
Second fit a new model classification over convolution features of SNRs.

Results

Third show results using full dataset of SNRs and only two features to illustrate a boundary decision.

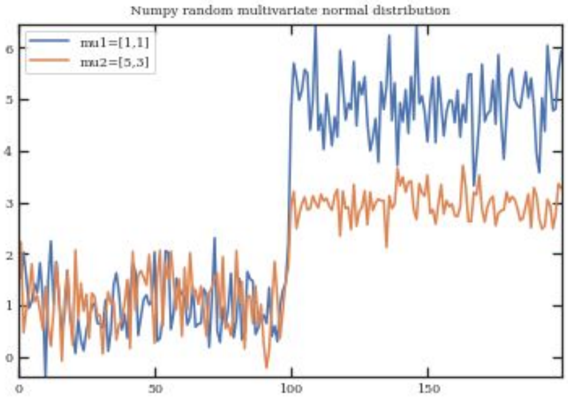
Reproduce

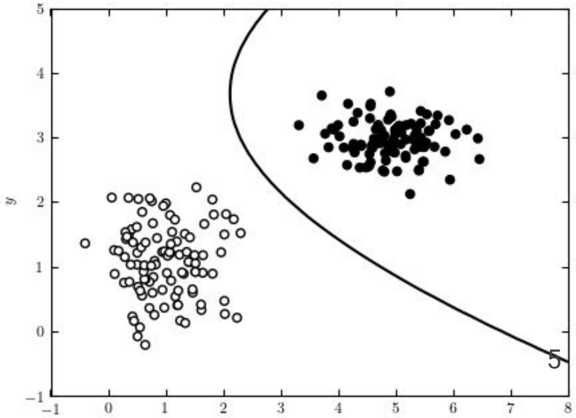
AstroML example at Chapter 9
Fig. 9.2



Gaussian Navie Bayes Classification

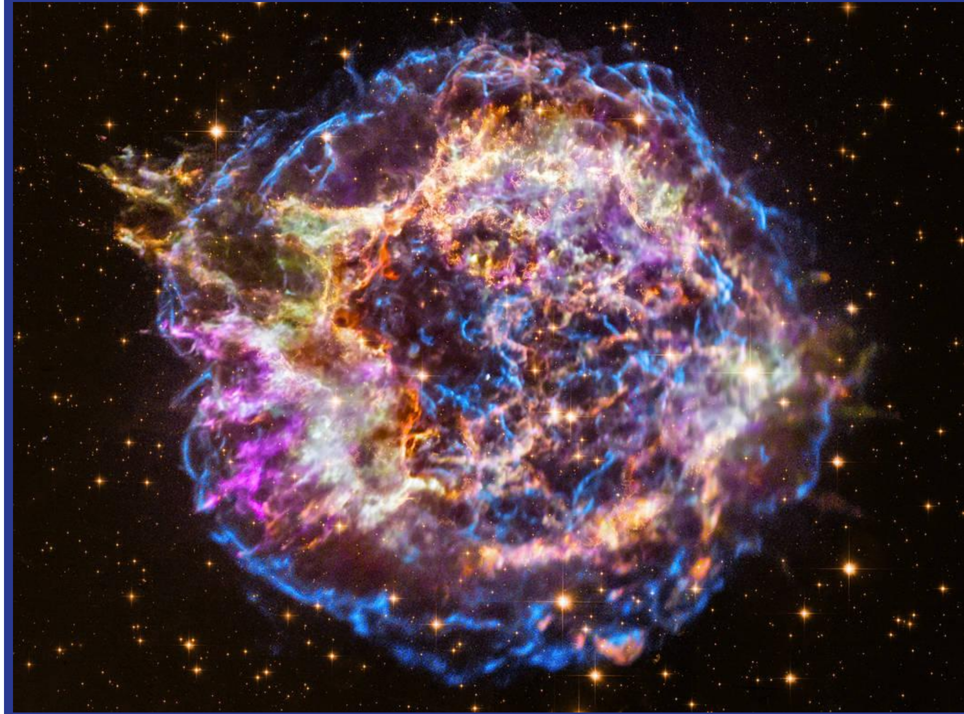
A decision boundary computed for a simple data set using Gaussian naive Bayes classification. The line shows the decision boundary, which corresponds to the curve where a new point has equal posterior probability of being part of each class. In such a simple case, it is possible to find a classification with perfect completeness and contamination. This is rarely the case in the real world.





Apply

Gaussian Naive Bayes Classification to SNRs



The supernova remnant Cassiopeia A.

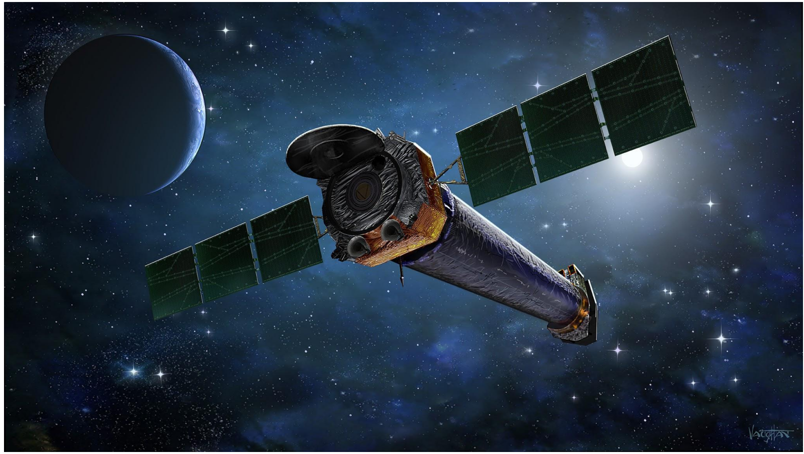
Source: Chandra Image Gallery available on https://www.nasa.gov/sites/default/files/styles/full_width_feature/public/thumbnails/image/firstlight_0.jpg

Chandra X-ray Observatory

Mission Overview

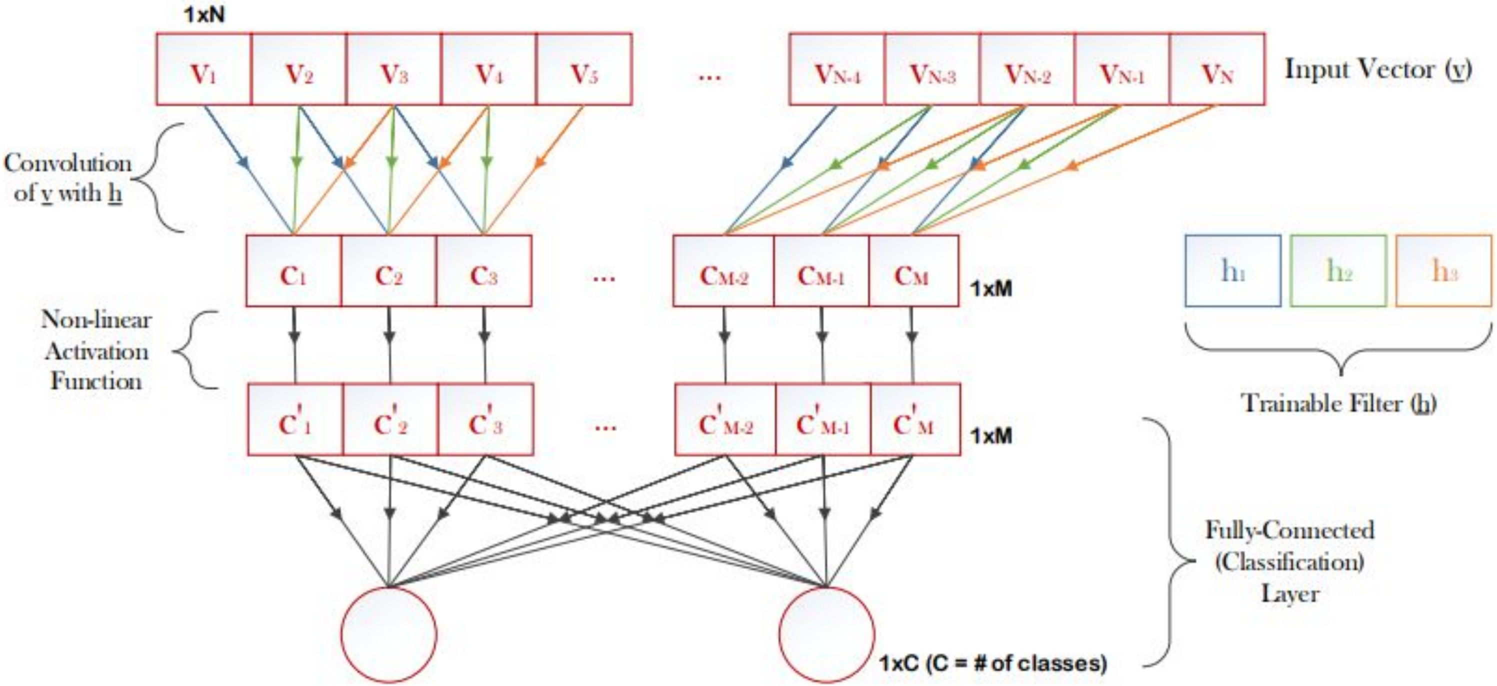
Launched: July 23, 1999

Goal: Detect X-ray emission from very hot regions of the Universe such as exploded stars, clusters of galaxies, and matter around black holes.



Source:
https://www.nasa.gov/mission_pages/chandra/main/index.html

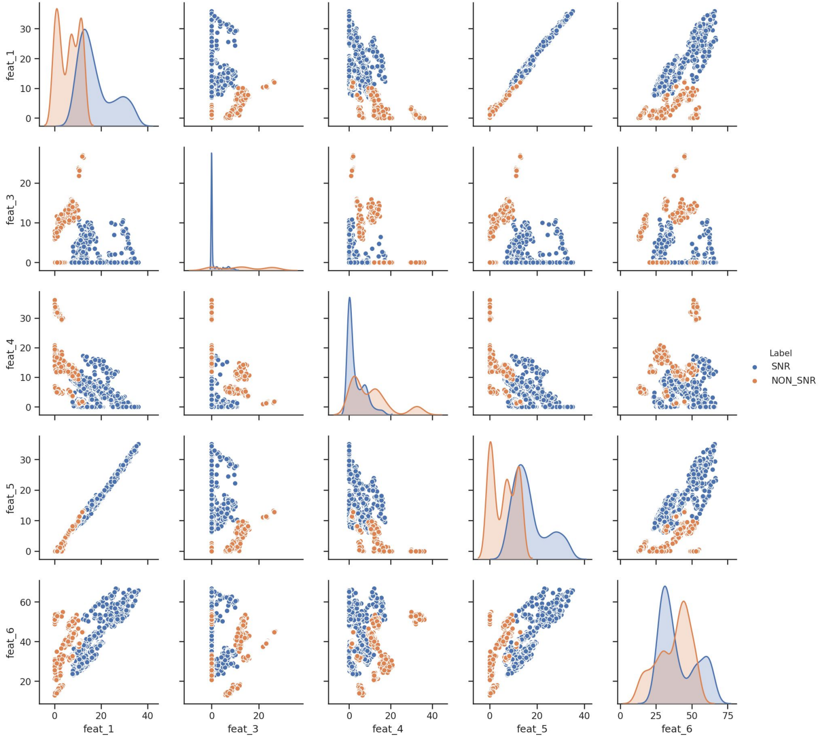
Convolution Neural Network



Source: Adapted from Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data available on https://arxiv.org/pdf/1809.09622.pdf

Convolution Neural Net Features

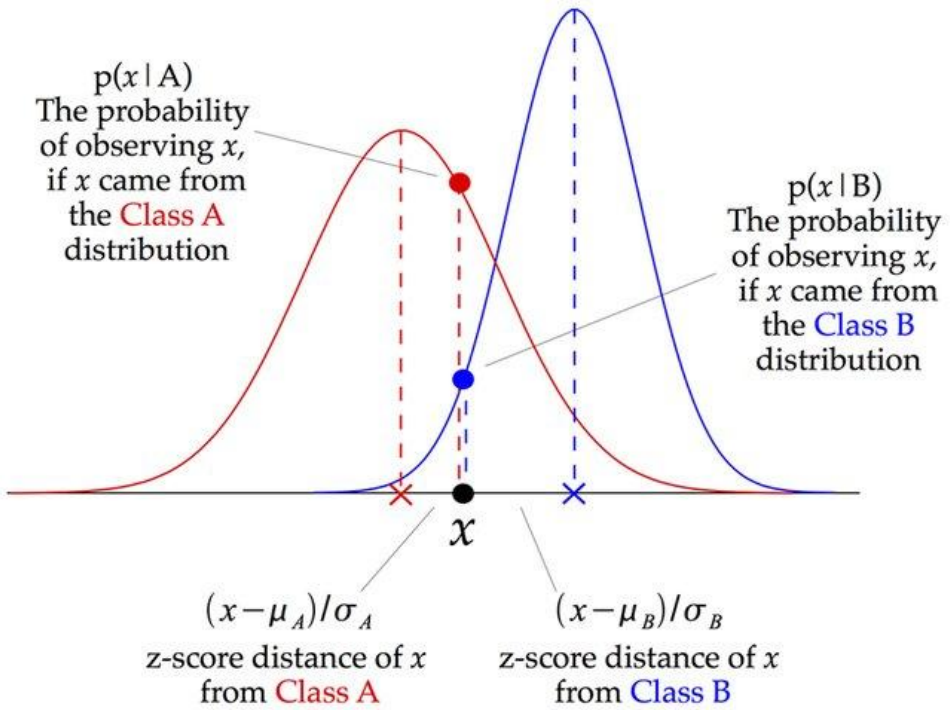
This dataset correspond to fifth layer of convolution neural network used to extract features from 30 images of SN1979C available in Chandra Supernova Remnant Catalog.[1]
Further detail please refer Arantes, Rosa and Guimarães (2019) at The Radio Universe: Prospects in Instrumentation and Data Science.



[1] Chandra Supernova Remnant Catalog available on https://hea-www.harvard.edu/ChandraSNR/index.html

Results

Gaussian Naive Bayes Classification



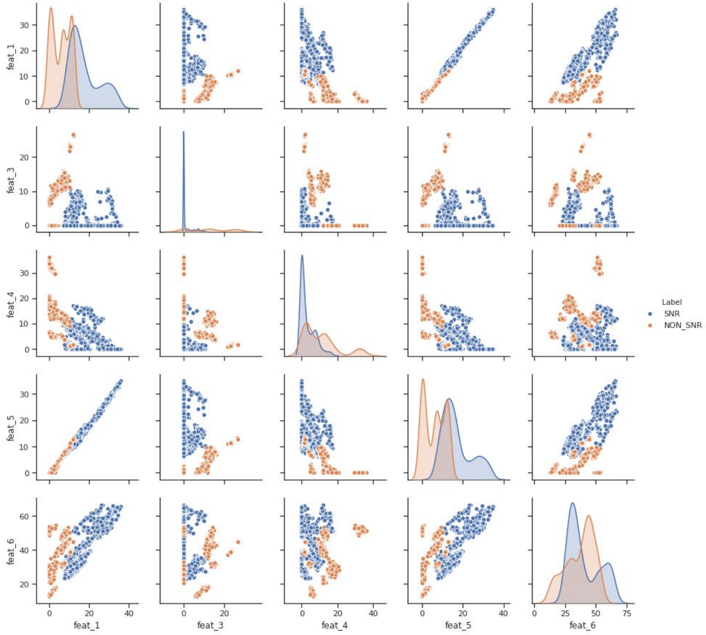
How Gaussian Naive Bayes Works.

Source: Adapted from Smoothness without Smoothing: Why Gaussian Naive Bayes Is Not Naive for Multi-Subject Searchlight Studies available on https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069566

Full dataset with SNR features

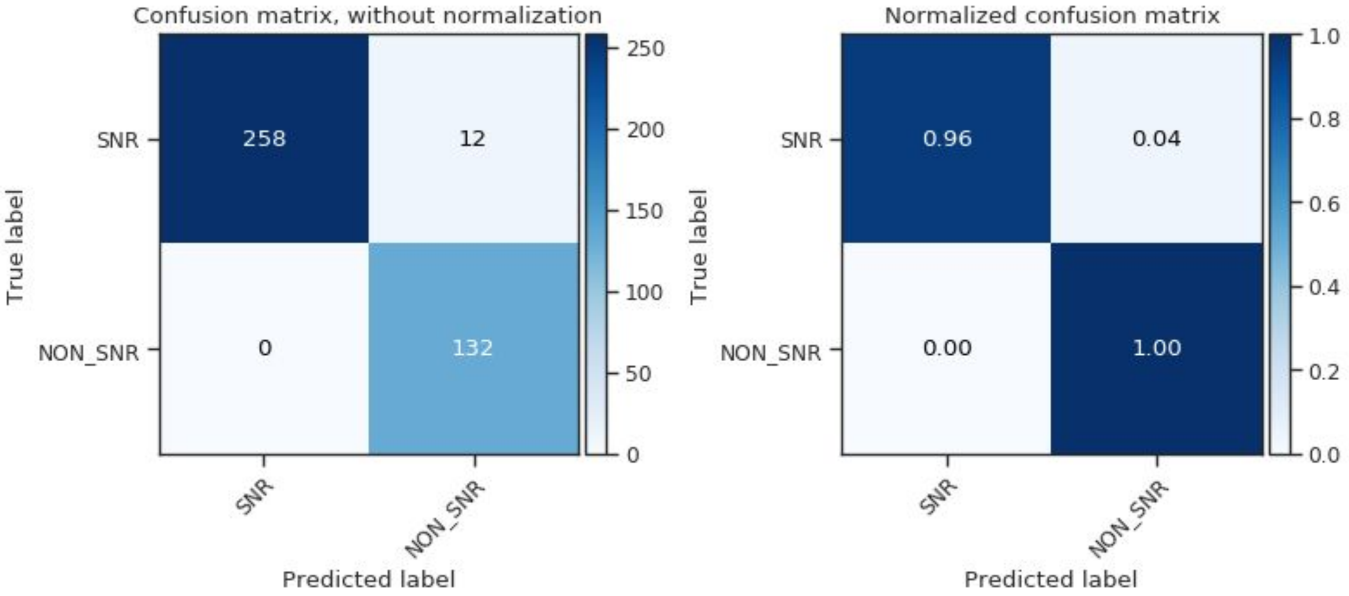
```
# Load SNR dataset
pathDataFrame = './topicos/'
df_SNR = pd.read_pickle(os.path.join(pathDataFrame,
'Data_Frame_SNR_features_7.pkl'))
print(df_SNR.head())
```

Tipo	feat_1	feat_2	feat_3	feat_4	feat_5	feat_6
1	16.432047	0.0	0.000000	0.000000	15.830933	32.354202
1	16.537399	0.0	0.000000	16.258198	13.950276	56.827934
2	0.136962	0.0	0.000000	17.344286	0.000000	29.395622
1	11.908845	0.0	1.177783	0.540163	12.111457	28.434286
2	12.066059	0.0	26.854000	1.761908	12.907033	44.984547

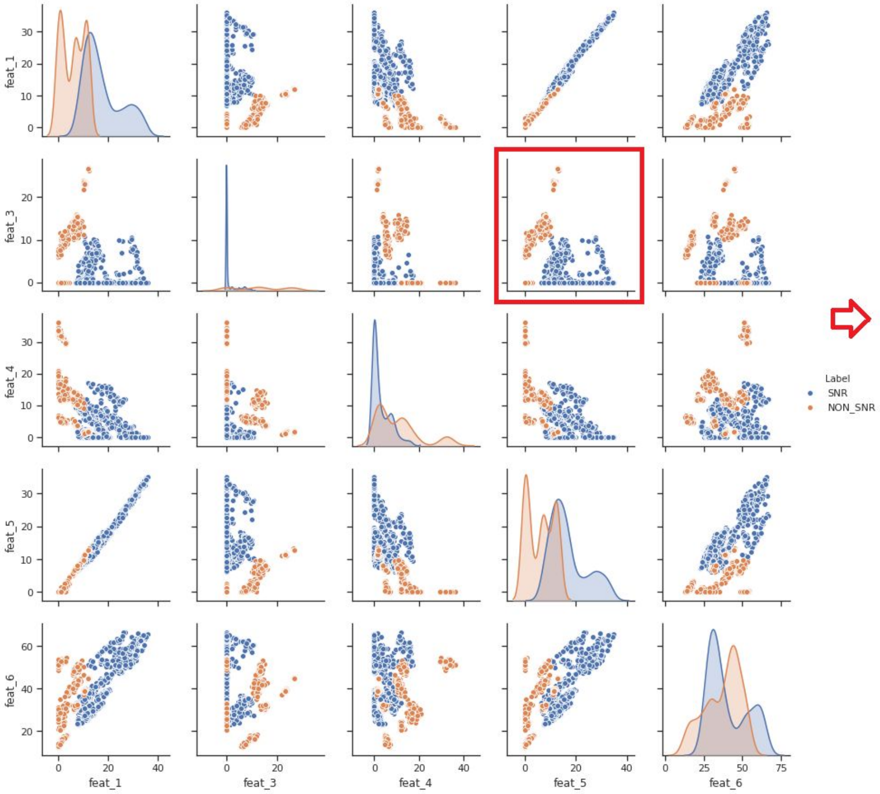


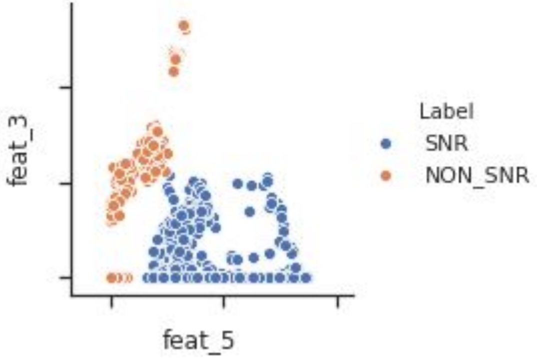
Result of full dataset

	precision	recall	f1-score	support
SNR	1.00	0.96	0.98	270
NON_SNR	0.92	1.00	0.96	132
accuracy			0.97	402



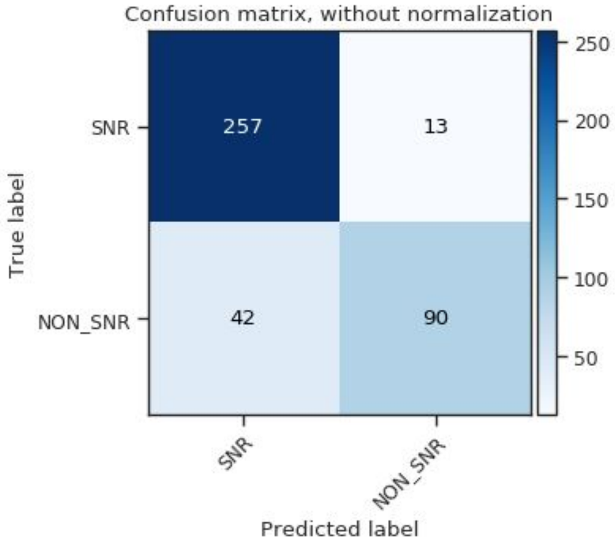
Revisit correlation matrix and choose features to ilustrate boundary decision

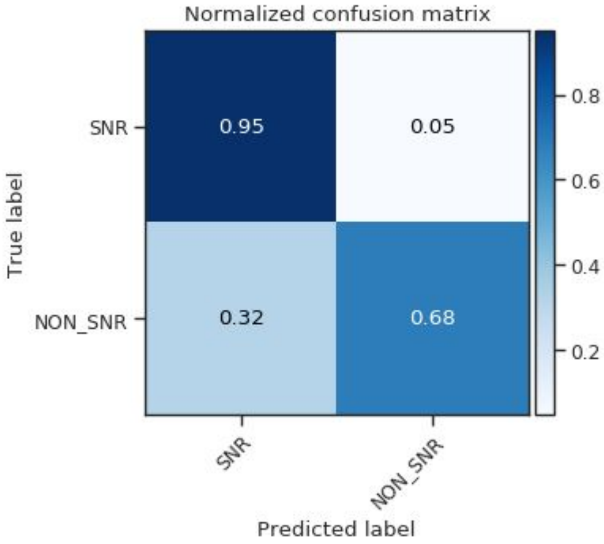




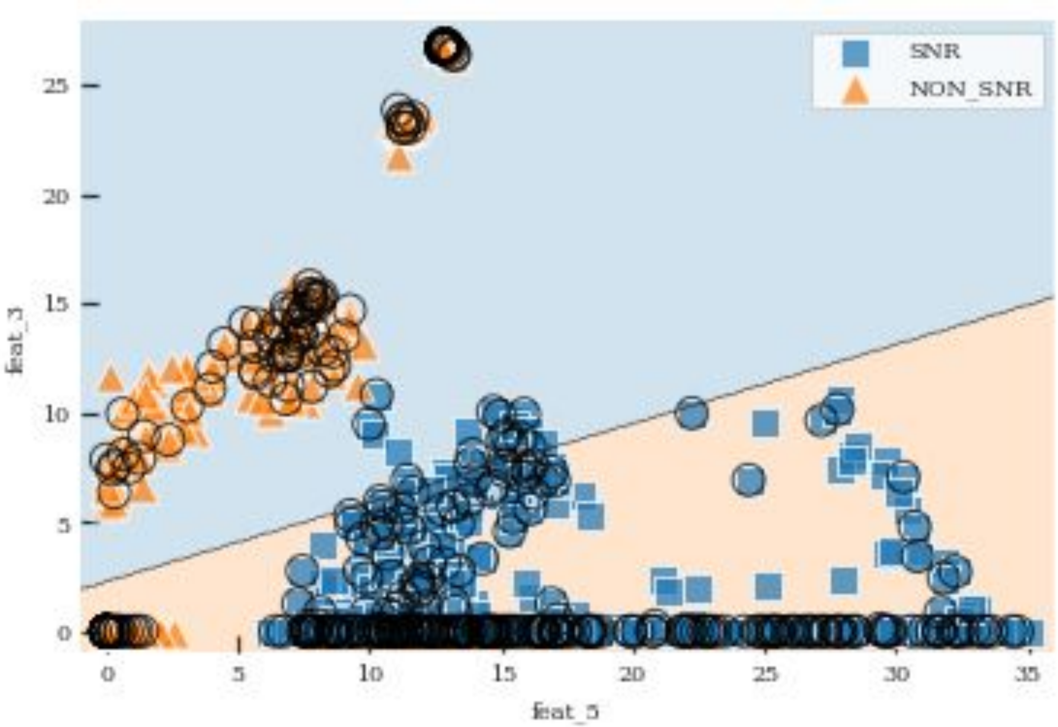
Result of partial dataset

	precision	recall	f1-score	support
SNR	0.86	0.95	0.90	270
NON_SNR	0.87	0.68	0.77	132
accuracy			0.86	402



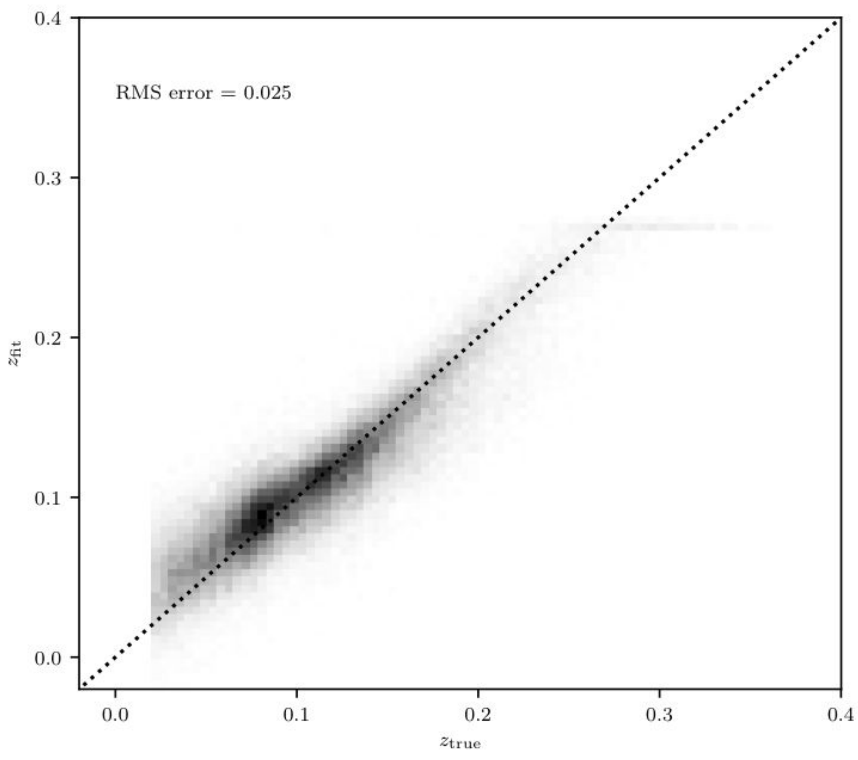


Result of partial dataset - decision boundary



Reproduce

AstroML example of correlation between measurement and predict redshifts

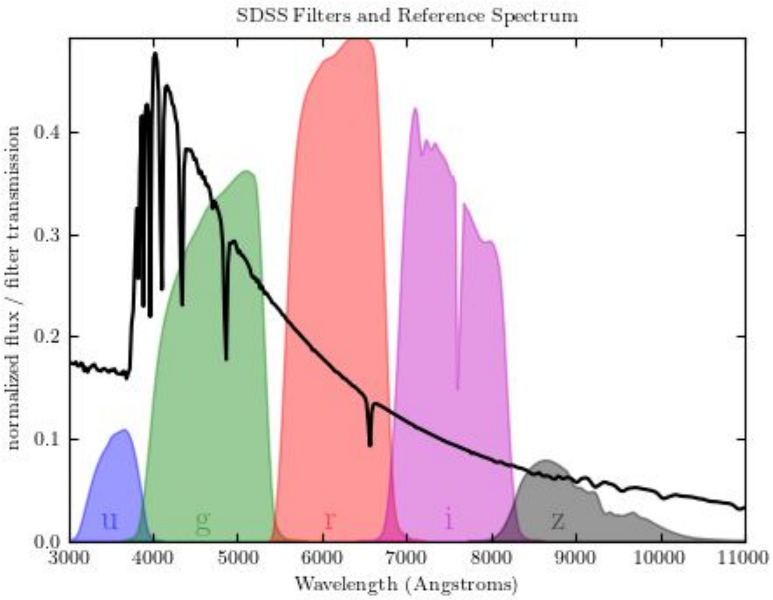


Dataset used to adjust MLP Neural Network

```
# Fetch and prepare the data
data = fetch_sdss_specgals()  [1]

# put magnitudes into array
# normalize to zero mean and unit variance for easier training
datanormed = np.zeros((len(data), 6), dtype=np.float32)
for i, band in enumerate(['u', 'g', 'r', 'i', 'z']):
    band = 'modelMag_' + band
    datanormed[:, i] = (data[band] - data[band].mean()) / data[band].std()

# put redshifts into array
datanormed[:, 5] = data['z']
```



[1] Loader for SDSS Galaxies with spectral information avaliable on https://www.astroml.org/modules/generated/astroML.datasets.fetch_sdss_specgals.html

Pytorch library

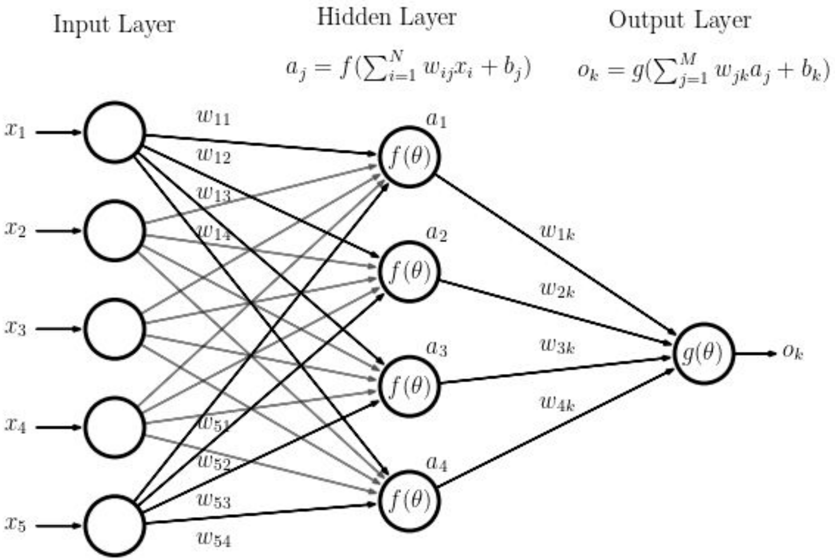
Pytorch is an open source machine learning library based on the Torch library (COLLOBERT; KAVUKCUOGLU; FARABET, 2011), used for applications such as computer vision and natural language processing. This library provides Tensors and Dynamic neural networks in Python with strong GPU acceleration.[1]

[1] Pytorch is available on https://pypi.org/project/torch

Build MLP Neural Network

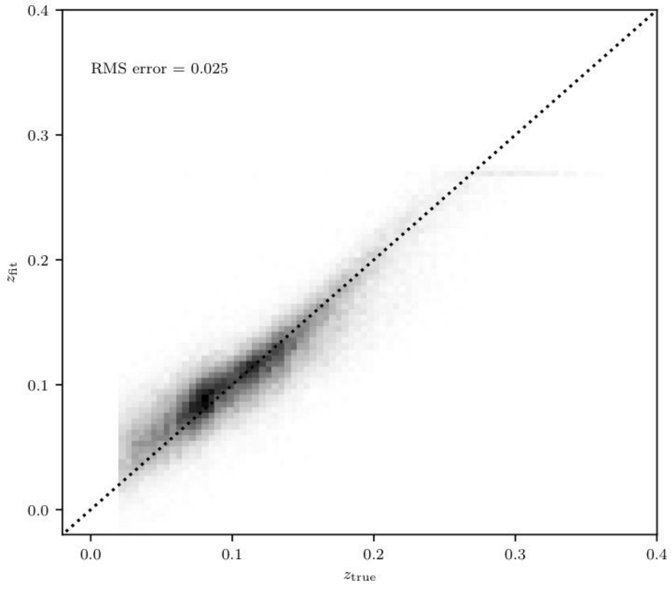
```
# define structure of neural net
class Net(nn.Module):
    def __init__(self, nhidden):
        super(Net, self).__init__()
        self.fc_h = nn.Linear(5, nhidden)
        self.fc_o = nn.Linear(nhidden, 1)

    def forward(self, x):
        h = F.relu(self.fc_h(x))
        z = self.fc_o(h)
        return z
...
model = Net(4)
criterion = torch.nn.MSELoss(reduction='sum')
optimizer = torch.optim.SGD(model.parameters(), lr=0.001)
scheduler = torch.optim.lr_scheduler.ReduceLROnPlateau(optimizer,
verbose=True, patience=5, threshold=1e-3)
```



Result MLP Neural Network

Epoch 0: train loss 1.169e-03 validation loss 7.723e-04
Epoch 10: train loss 5.723e-04 validation loss 5.878e-04
Epoch 20: train loss 5.605e-04 validation loss 5.737e-04
Epoch 30: train loss 5.537e-04 validation loss 5.704e-04
Epoch 40: train loss 5.490e-04 validation loss 5.678e-04
Epoch 48: reducing learning rate of group 0 to 1.0000e-04.
Epoch 50: train loss 5.439e-04 validation loss 5.632e-04
Epoch 55: reducing learning rate of group 0 to 1.0000e-05.
Finished training



References

ARANTES FILHO, L. R.; ROSA, R. R.; GUIMARÃES, L. N. F. Deep Learning Approach to Retrieve Image Features of Radio Supernovae Remnants. The Radio Universe: Prospects in Instrumentation and Data Science. Anais...Sao Jose dos Campos: 2019 Disponível em: <http://www.inpe.br/radiouniverse/2019/>

COLLOBERT, Ronan; KAVUKCUOGLU, Koray; FARABET, Clément. Torch7: A matlab-like environment for machine learning. In: BigLearn, NIPS workshop. 2011.