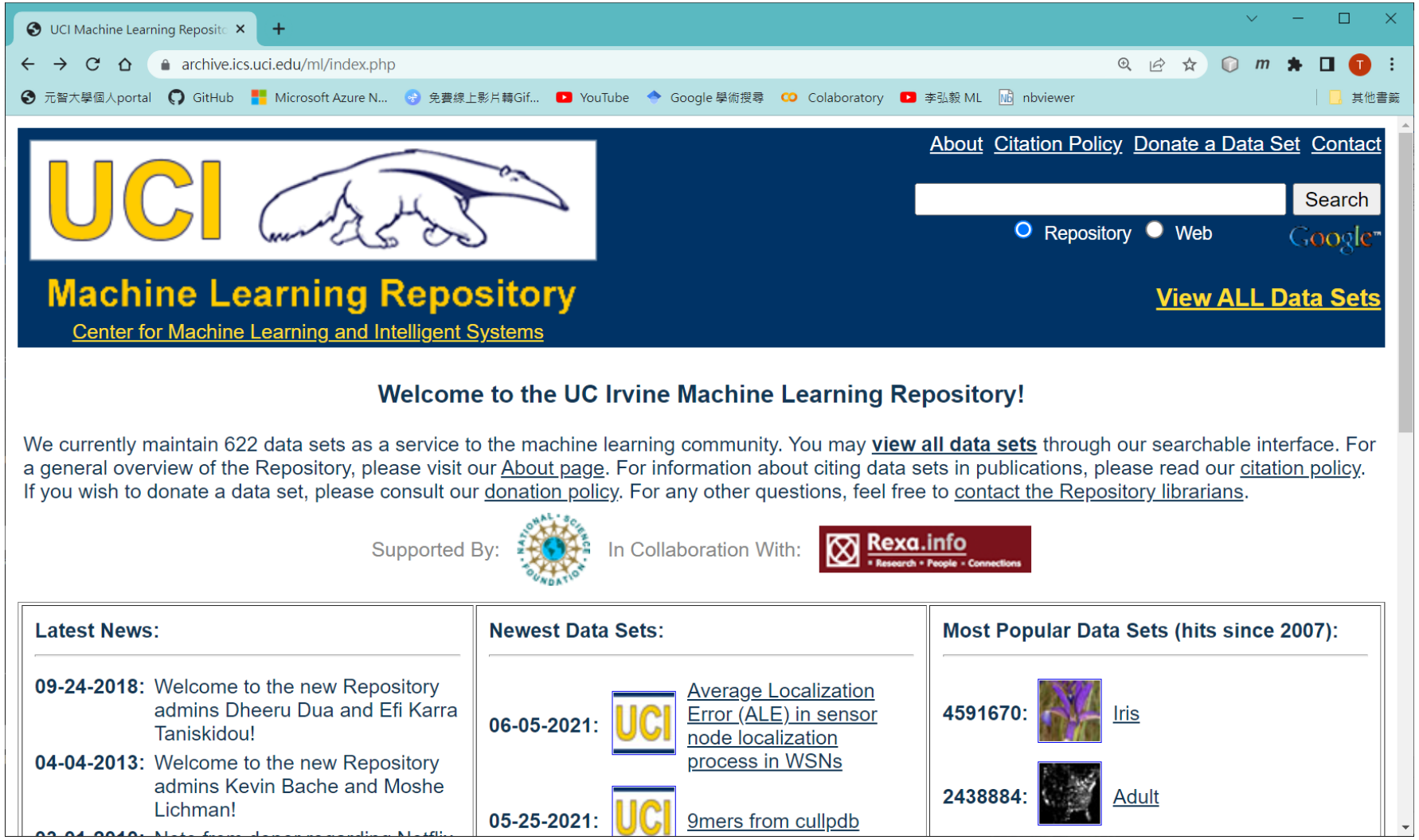
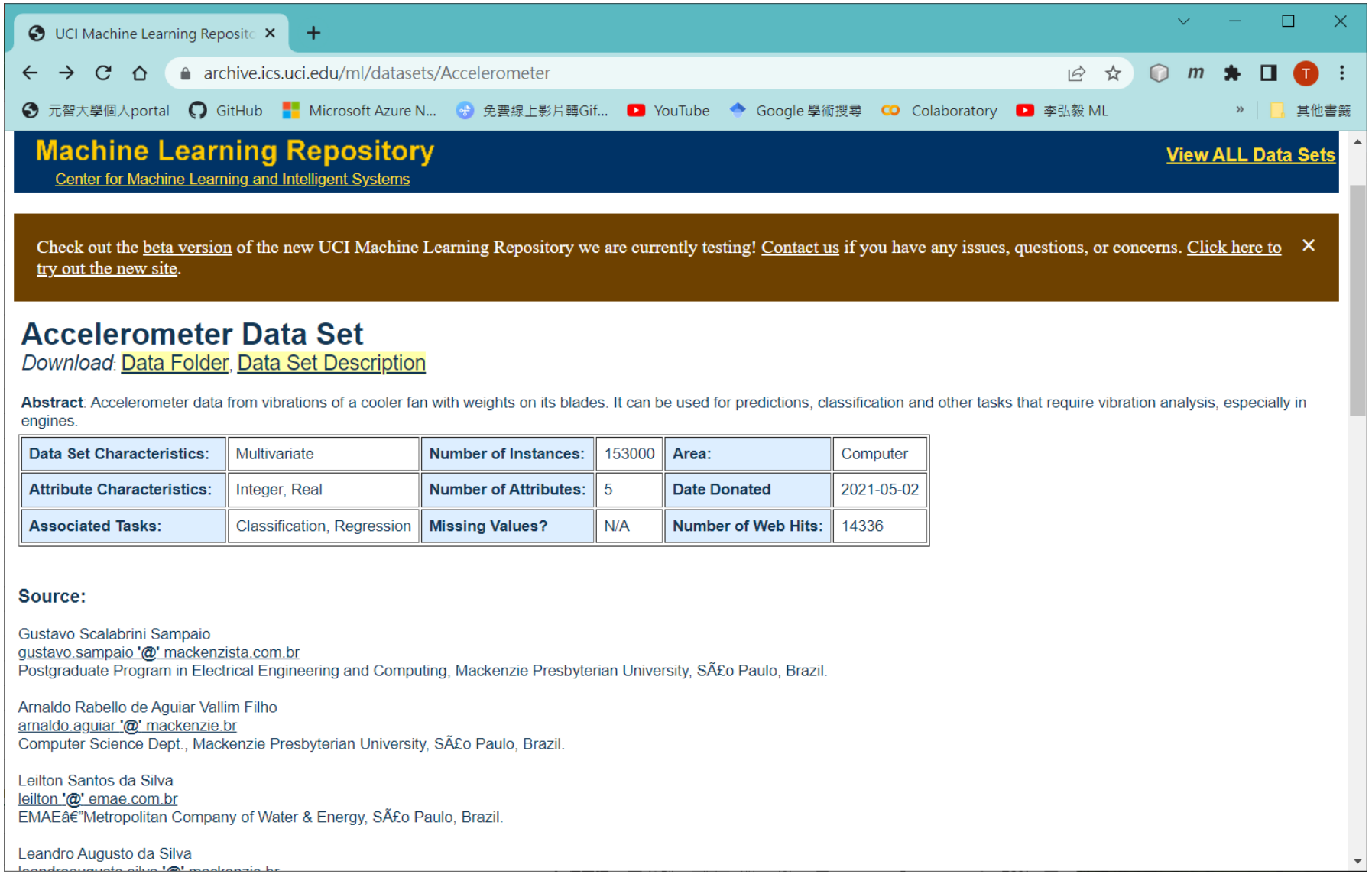
Find out a data set for your HW

Google search: UCI Machine Learning Repository



Example: Accelerometer data set



UCI Machine Learning Reposit

archive.ics.uci.edu/ml/datasets/Accelerometer

元智大學個人portal GitHub Microsoft Azure N... 免費線上影片轉Gif... YouTube Google 學術搜尋 Colaboratory 李弘毅 ML 其他書籤

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Check out the beta version of the new UCI Machine Learning Repository we are currently testing! Contact us if you have any issues, questions, or concerns. Click here to try out the new site.

Accelerometer Data Set

Download: Data Folder, Data Set Description

Abstract: Accelerometer data from vibrations of a cooler fan with weights on its blades. It can be used for predictions, classification and other tasks that require vibration analysis, especially in engines.

Data Set Characteristics:	Multivariate	**Number of Instances:**	153000	**Area:**	Computer
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	5	**Date Donated**	2021-05-02
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	14336

Source:

Gustavo Scalabrini Sampaio
gustavo.sampaio '@' mackenzista.com.br
Postgraduate Program in Electrical Engineering and Computing, Mackenzie Presbyterian University, SÃ£o Paulo, Brazil.

Arnaldo Rabello de Aguiar Vallim Filho
arnaldo.aguiar '@' mackenzie.br
Computer Science Dept., Mackenzie Presbyterian University, SÃ£o Paulo, Brazil.

Leilton Santos da Silva
leilton '@' emae.com.br
EMAEâ€"Metropolitan Company of Water & Energy, SÃ£o Paulo, Brazil.

Leandro Augusto da Silva

Feature scaling

Feature scaling - Process your data in Excel

wconfid	pctid	x	y	z
1	20	1.004	0.09	-0.125
1	20	1.004	-0.043	-0.125
1	20	0.969	0.09	-0.121
1	20	0.973	-0.012	-0.137
1	20	1	-0.016	-0.121
1	20	0.961	0.082	-0.121
1	20	0.973	-0.055	-0.109
1	20	1	0.012	-0.133
1	20	0.969	-0.102	-0.141
1	20	0.973	-0.059	-0.125
1	20	1.012	0.043	-0.133
1	20	0.996	-0.109	-0.148
1	20	0.988	-0.02	-0.125
1	20	1.012	0.043	-0.129
1	20	0.996	-0.09	-0.152
1	20	0.965	-0.102	-0.117
1	20	1.004	0.055	-0.121
1	20	0.988	-0.059	-0.141
1	20	0.969	-0.086	-0.117

y format is different in regression vs classification

2.3. Regression HW.ipynb

```
dfY = df[['Appliances']]
```

```
In [15]: print(numpyX.shape, numpyY.shape)
```

```
(19735, 25) (19735, 1)
```

4. Classifier.ipynb

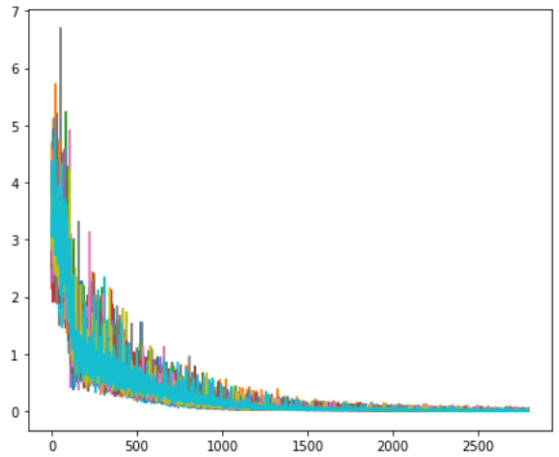
```
[5]: #convert data to numpy array
numpyX = np.array(lstX)
numpyY = np.array(lstY)
print(numpyX.shape, numpyY.shape)
```

```
(2200, 7) (2200,)
```

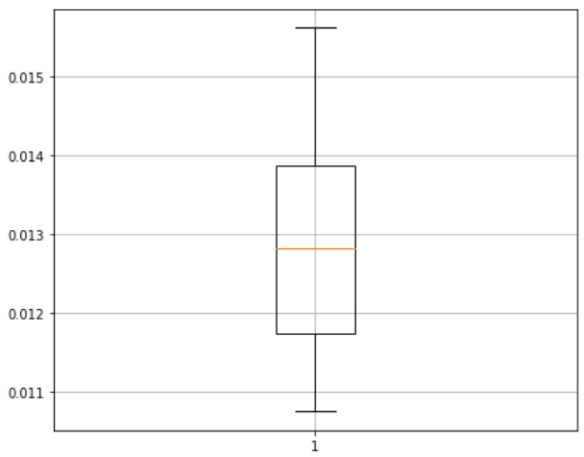
You need to modify: dfY=df['...']

Loss plot reading is different

2.1. Regression.ipynb

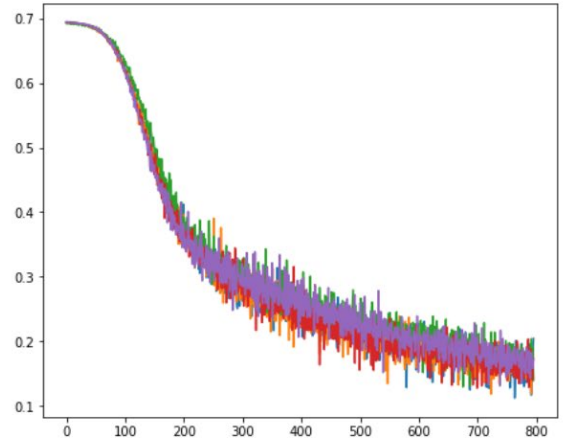


MSE of training data

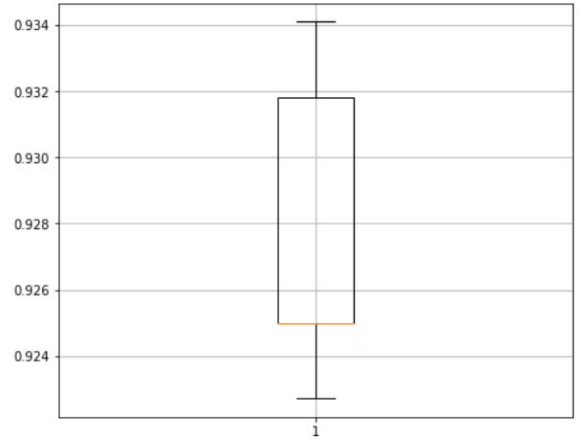


MSE of test data

4. Classifier.ipynb



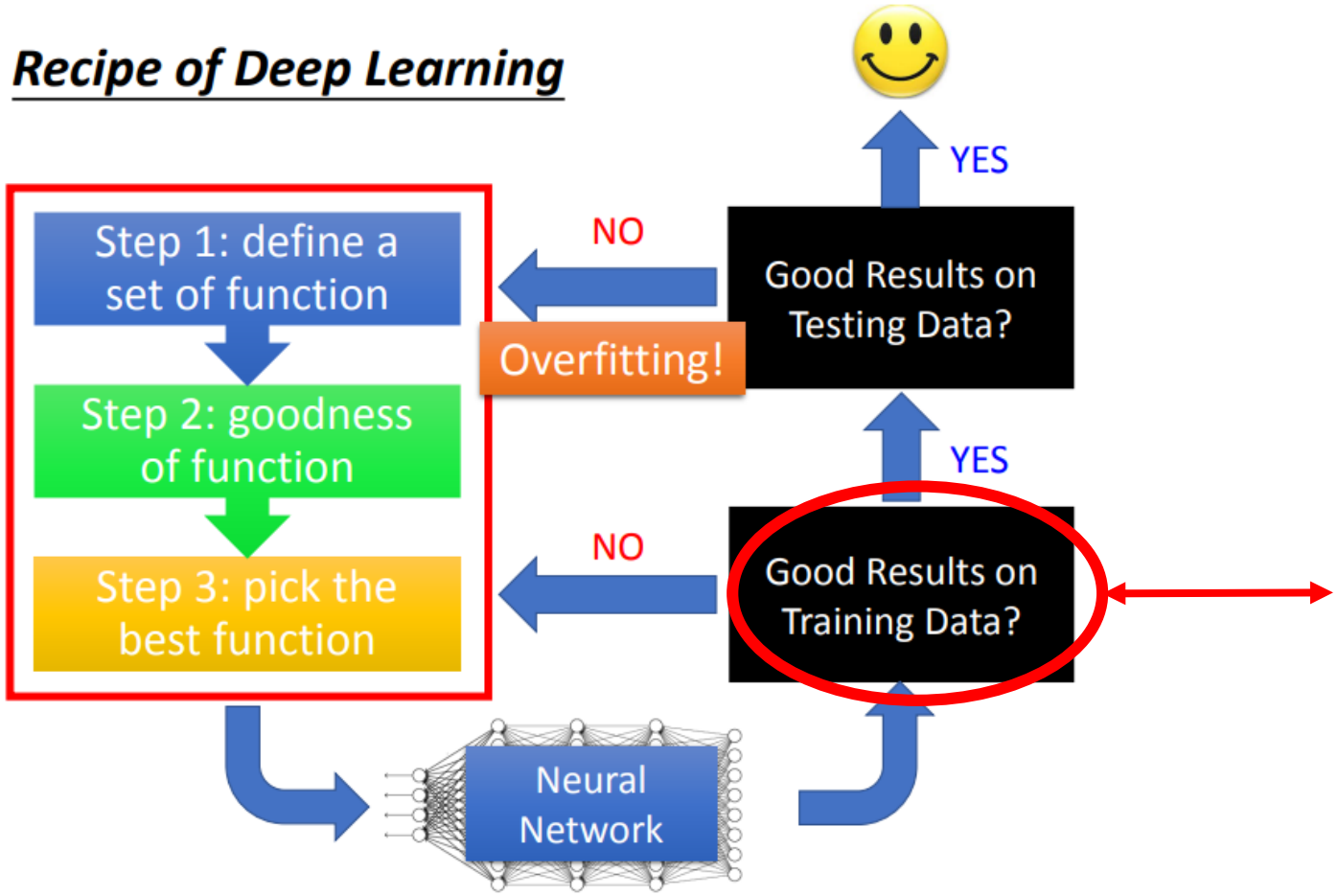
CE of training data

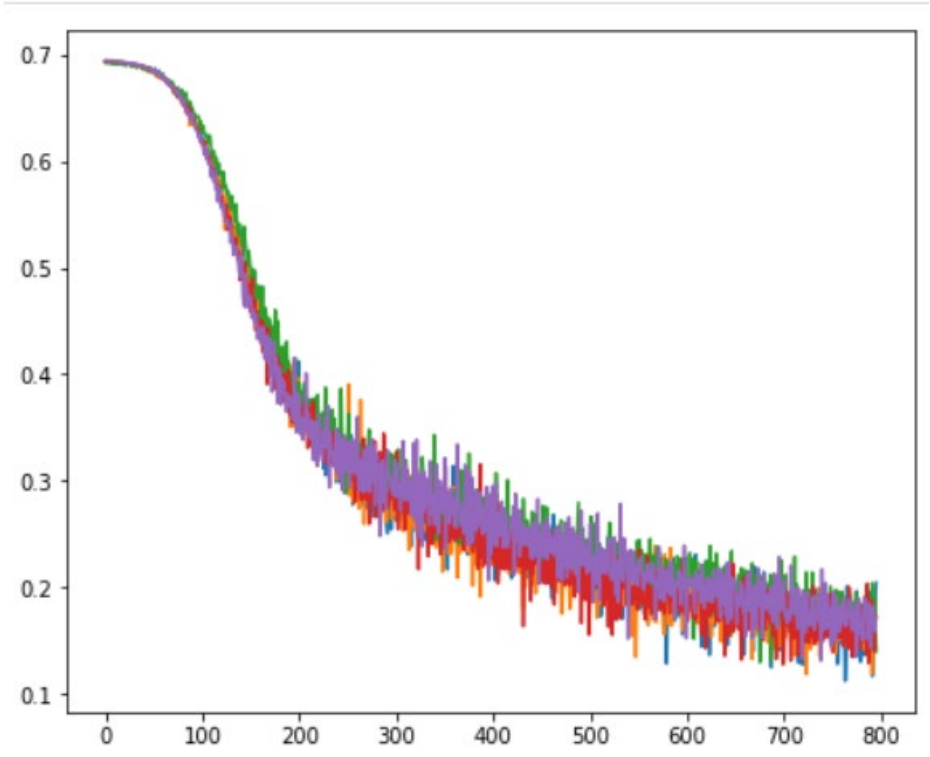


Accuracy of test data

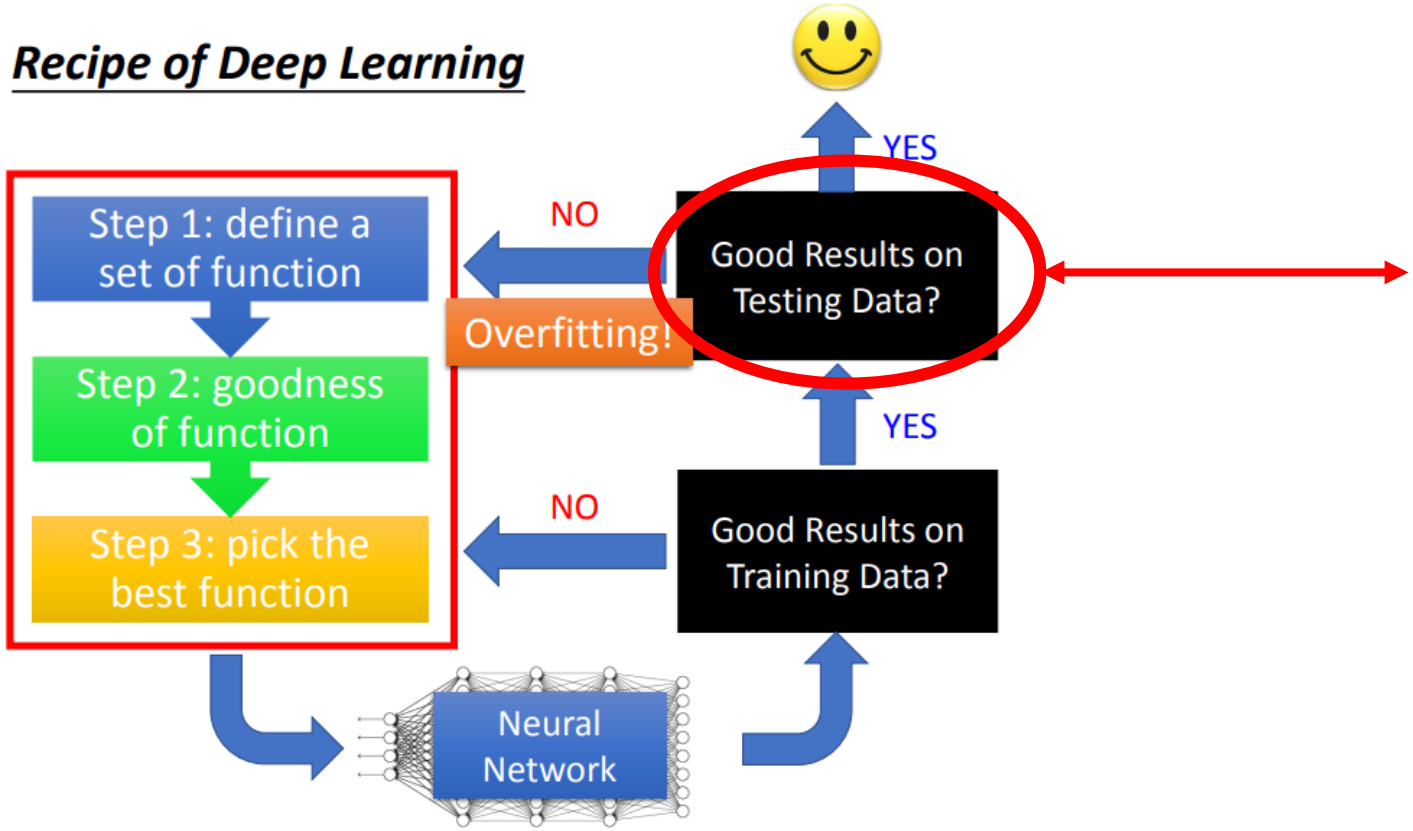
How to present misclassified test data?

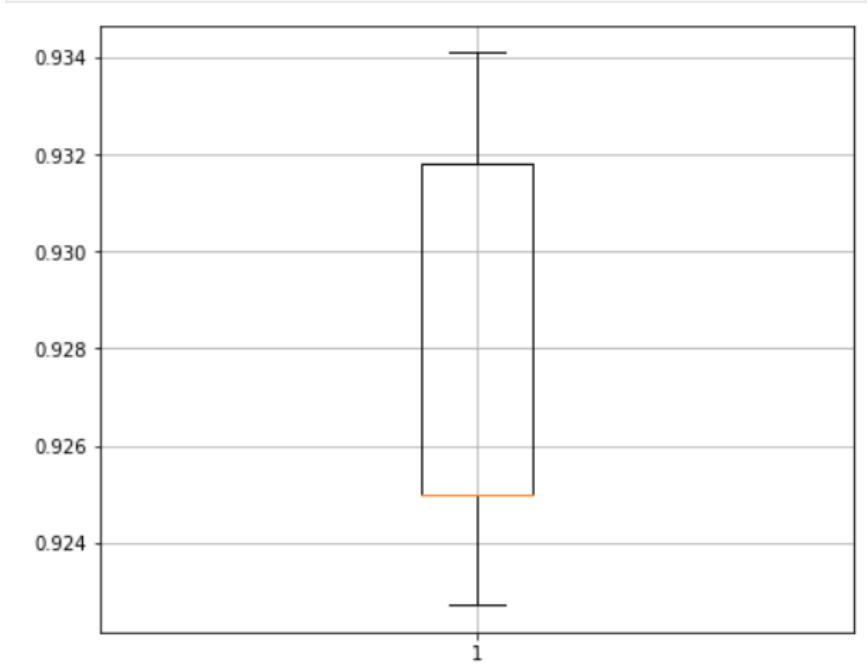
How is the training performance?





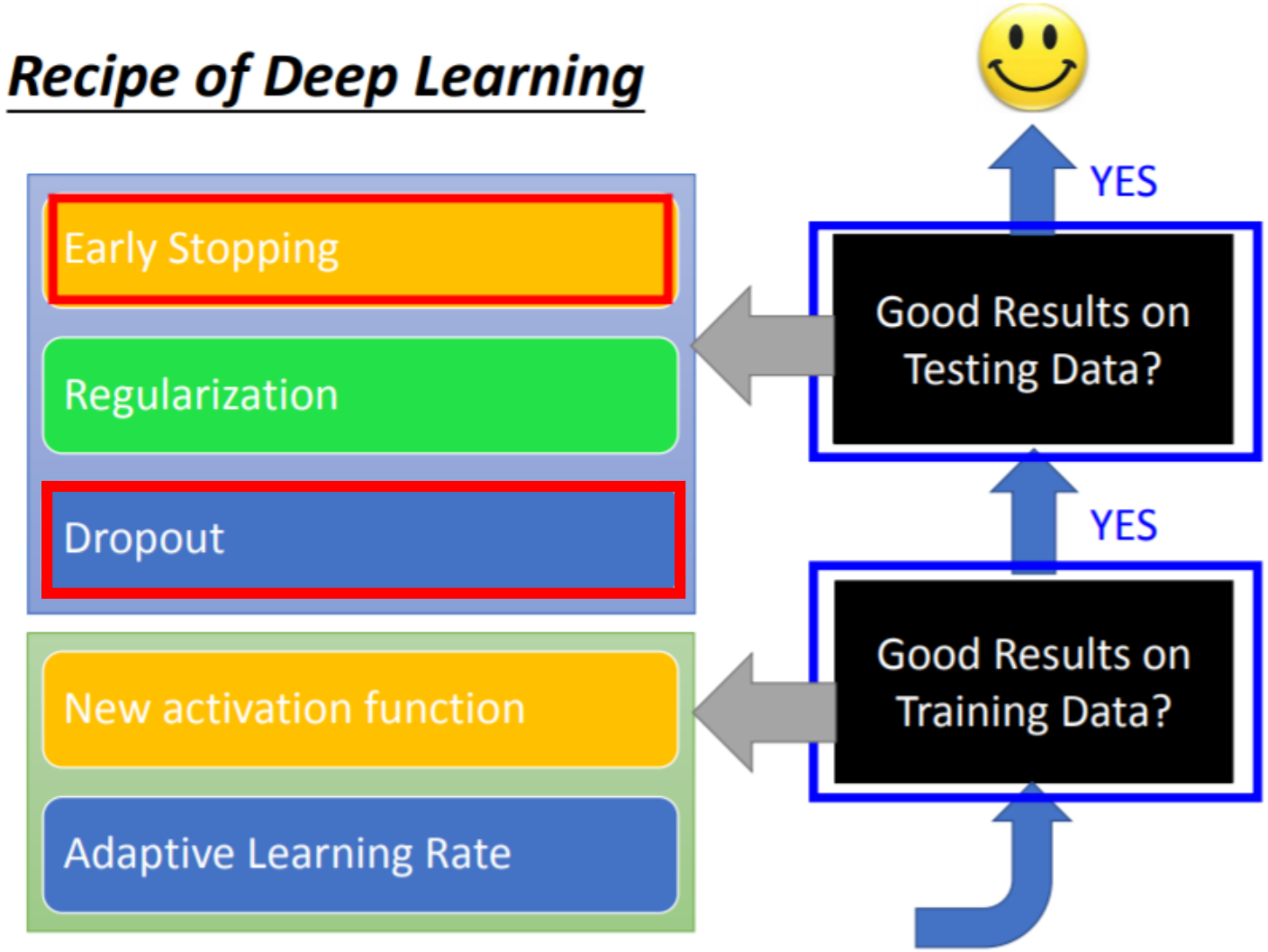
How is the test performance?





Accuracy only measures correctly classified test data.
How to measure misclassified test data?

Solve overfitting



Reference: 李弘毅 ML Lecture 9-1 https://youtu.be/xki61j7z-30