Machine Learning

FALL DETECTOR

A life well lived

Falls are leading cause of injury and death in older adults

According to the U.S. Centers for Disease Control and Prevention:

- One in four Americans aged 65+ falls each year.
- Every 11 seconds, an older adult is treated in the emergency room for a fall; every 19 minutes, an older adult dies from a fall.
- Falls are the leading cause of fatal injury and the most common cause of nonfatal trauma-related hospital admissions among older adults.
- Falls result in more than 2.8 million injuries treated in emergency departments annually, including over 800,000 hospitalizations and more than 27,000 deaths.

Fall Detection and Movement Tracking

- Kaggle : Fall detection data from China
- A system designed to fulfill the need for a wearable device to collect data for fall and near-fall analysis.
- Four different fall trajectories (forward, backward, left and right), three normal activities (standing, walking and lying down) and near-fall situations are identified and detected.
- Fourteen volunteers perform a standardized set of movements including 20 voluntary falls and 16 activities of daily living (ADLs), resulting in a large dataset with 2520 trials.
-

Machine Learning Results

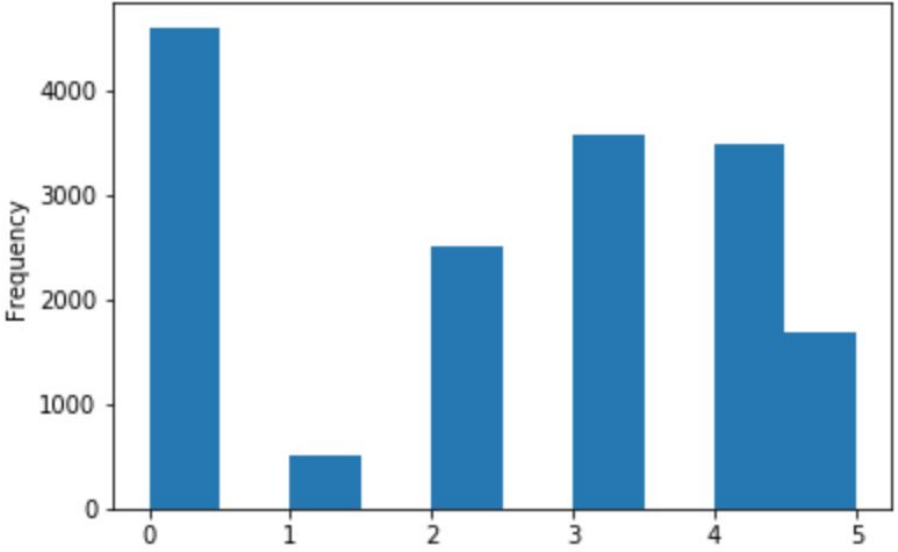
I successfully distinguish falls from ADLs using six machine learning techniques (classifiers): the k-nearest neighbor (k-NN) classifier, support vector machines (SVM), Bayesian decision making (BDM), Regression, and artificial neural networks (ANNs). I compare the performance and the computational complexity of the classifiers and achieve the best results with the **k-NN** classifier and **ANNs**, with accuracy all above **83%**. These classifiers also have acceptable computational requirements for training and testing. My approach would be applicable in real-world scenarios where data records of indeterminate length, containing multiple activities in sequence, are recorded.

About the dataset

- Fall detection data set of Chinese hospitals of old age patients.
- 0- Standing 1- Walking 2- Sitting 3- Falling 4- Cramps 5- Running
- Columns:
- ACTIVITY : activity classification
- TIME : monitoring time
- SL: sugar level
- EEG: EEG monitoring rate
- BP: Blood pressure
- HR: Heart beat rate
- CIRCLUATION: Blood circulation
-

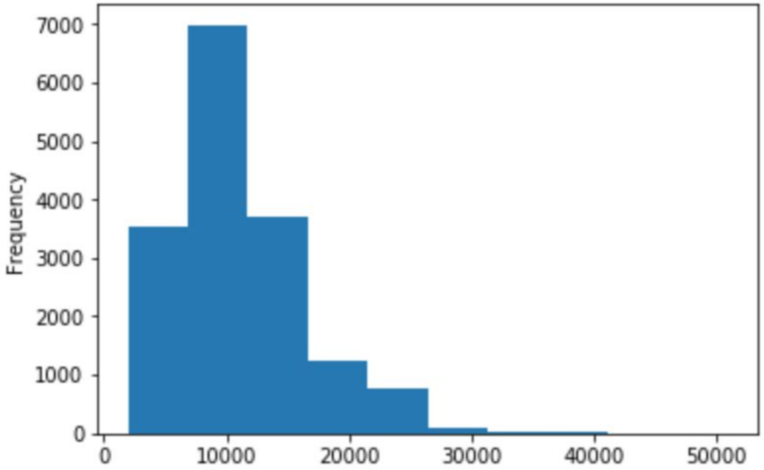
Visualize the data

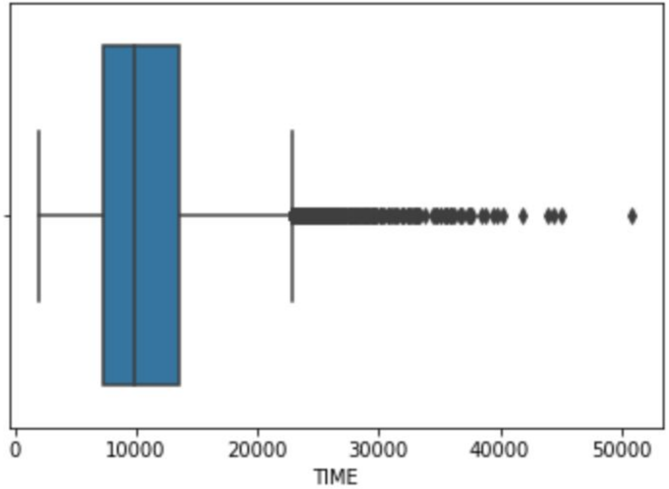
Activity distribution



Visualize the data

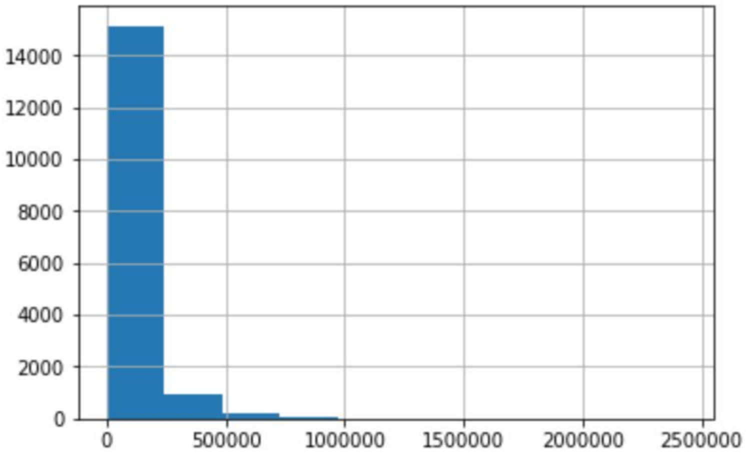
TIME Distribution

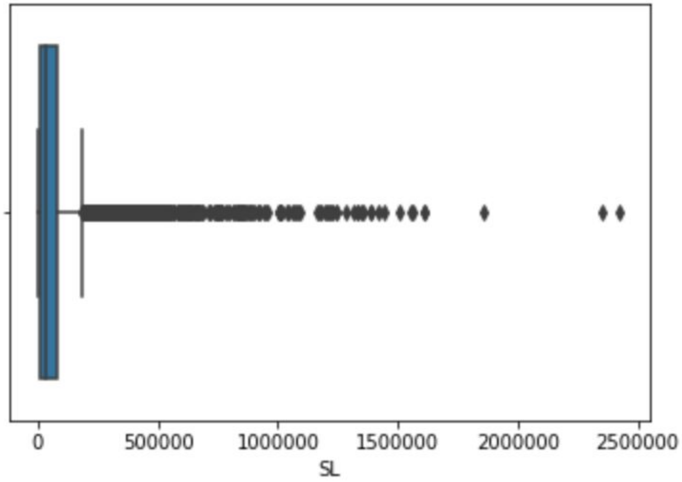




Visualize the data

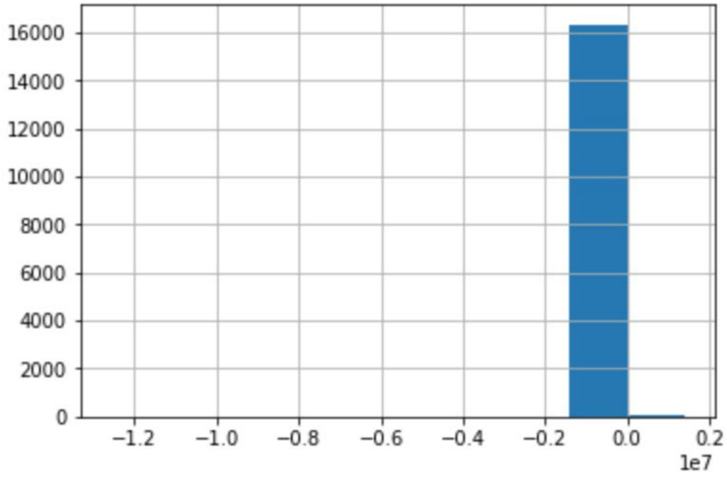
Sugar Level Distribution

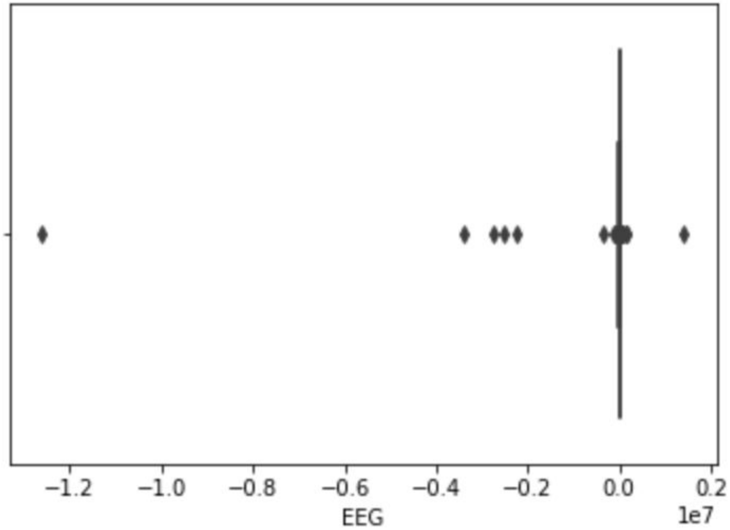




Visualize the data

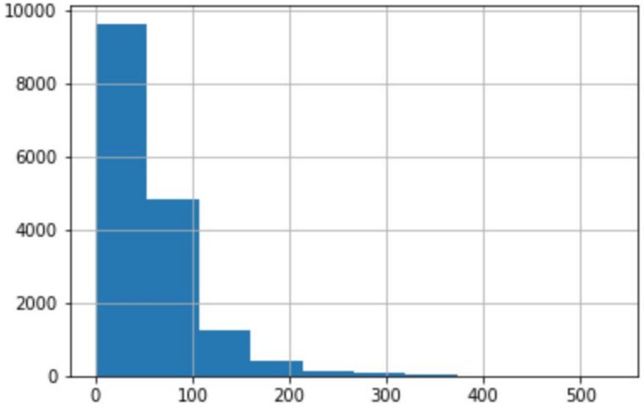
EEG Distribution

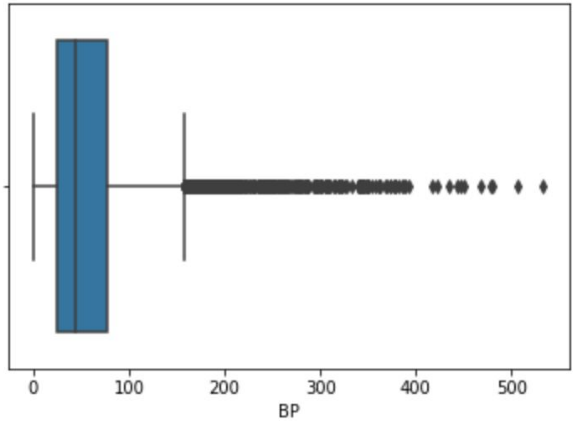




Visualize the data

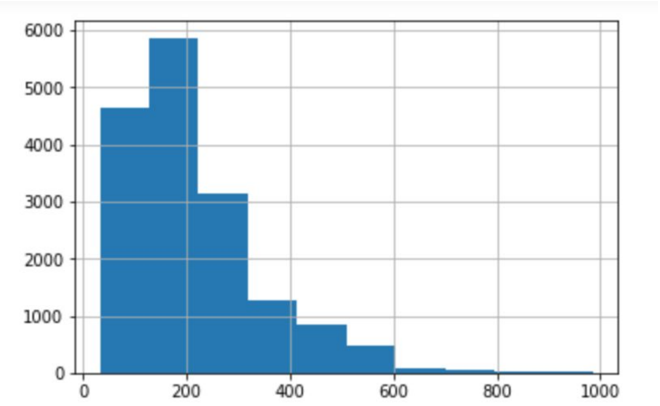
Blood Pressure Distribution

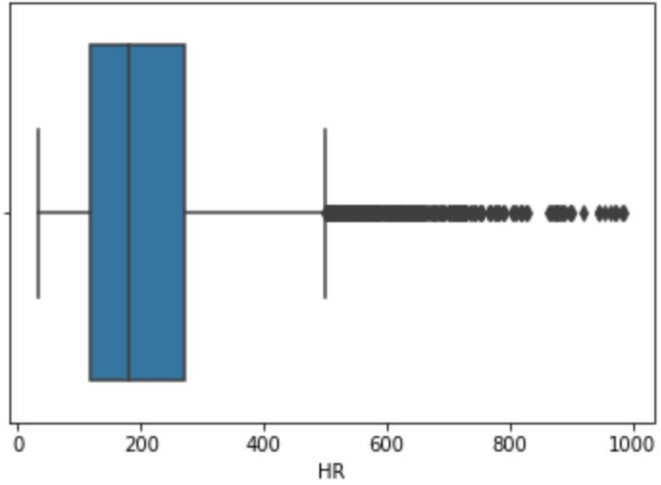




Visualize the data

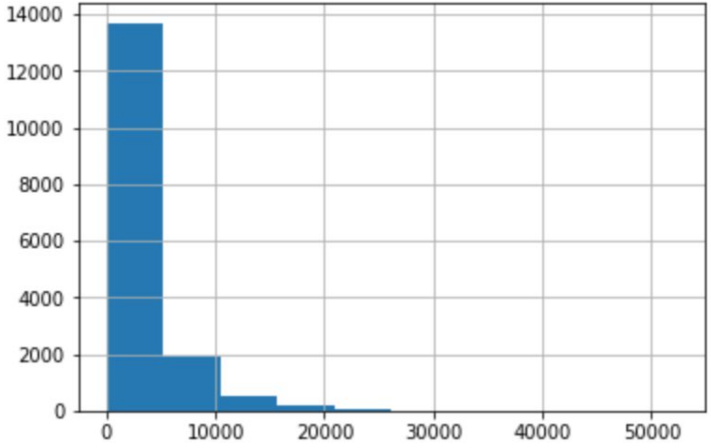
Heart Rate Distribution

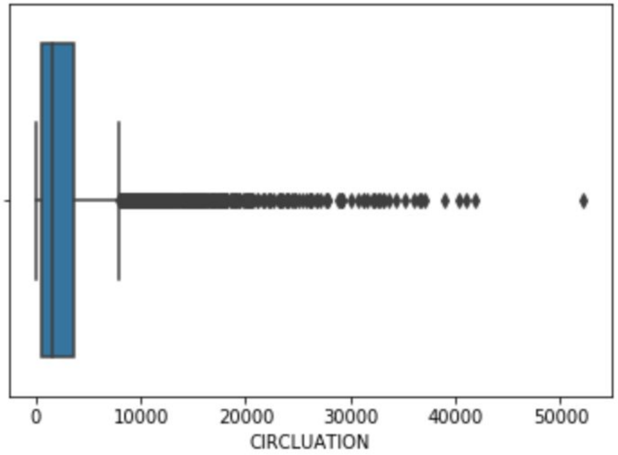




Visualize the data

Blood Circulation Distribution





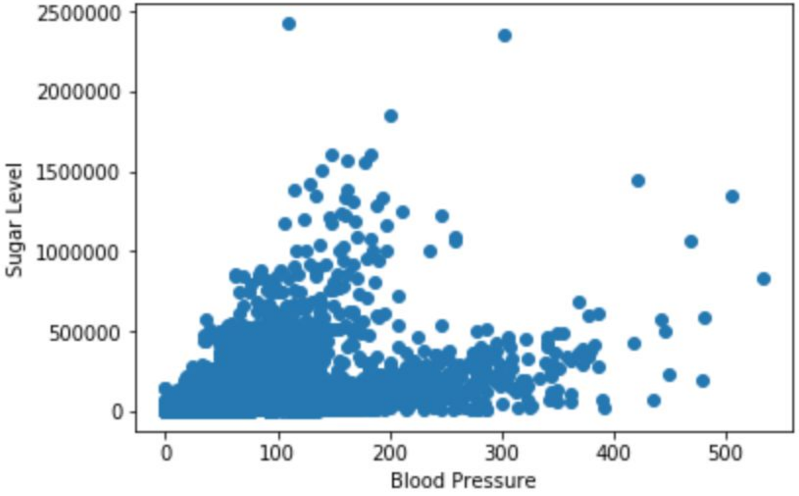
Remove Outliers

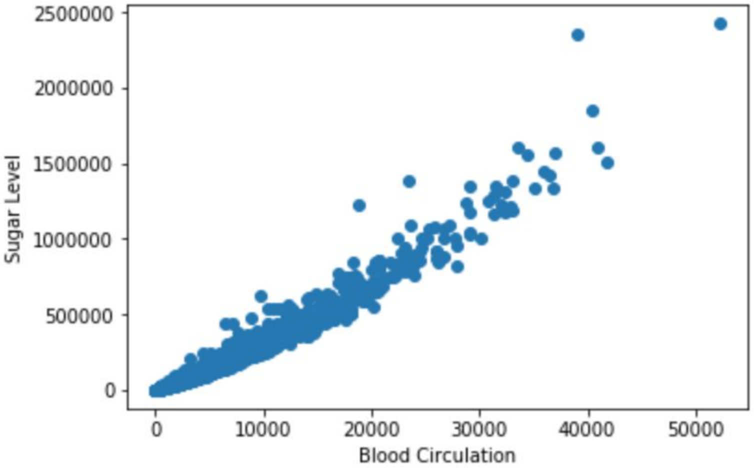
- Domain knowledge
- Visualize the data
- Use Z-score

Domain Knowledge

- Since the blood pressure machine would only measure up to 300, so all the BP>300 are false data.

Visualize the data





Z-Score method

- Data point that falls outside of 1.5 times of an interquartile range above the 3rd quartile and below the 1st quartile
- Data point that falls outside of 3 standard deviations. we can use a z score and if the z score falls outside of 2 standard deviation

The clean data

	ACTIVITY	TIME	SL	EEG	BP	HR	CIRCLUATION	FALL
0	3	4722.92	4019.64	-1600.00	13	79	317	1
1	2	4059.12	2191.03	-1146.08	20	54	165	0
2	2	4773.56	2787.99	-1263.38	46	67	224	0
3	4	8271.27	9545.98	-2848.93	26	138	554	0
4	4	7102.16	14148.80	-2381.15	85	120	809	0

Feature Scaling

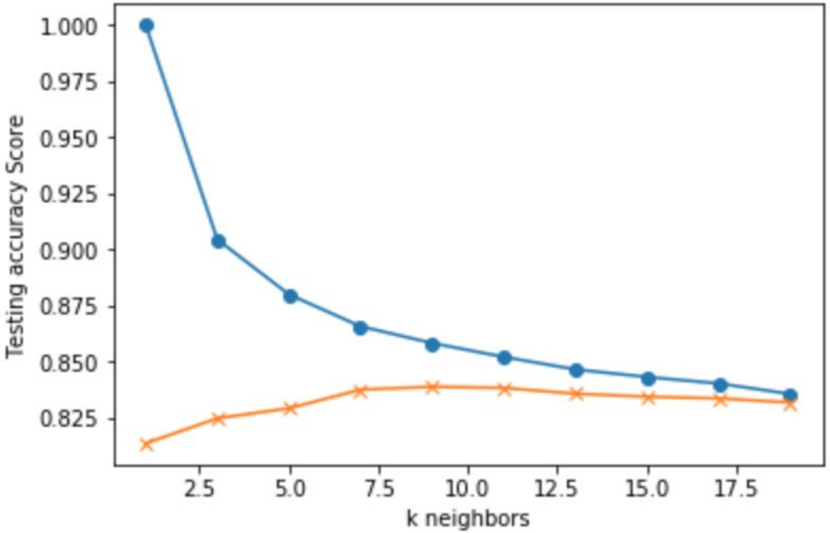
- Min- max scaling
- Standard scaling
- Robust scaler
- Quantile transfer

Decision Tree

- 0.23256160901784084, 'SL'
- 0.19585700909528694, 'EEG'
- 0.18456627794490044, 'TIME'
- 0.15784464096466386, 'BP'
- 0.11517915295809109, 'HR'
- 0.11399131001921678, 'CIRCLUATION'

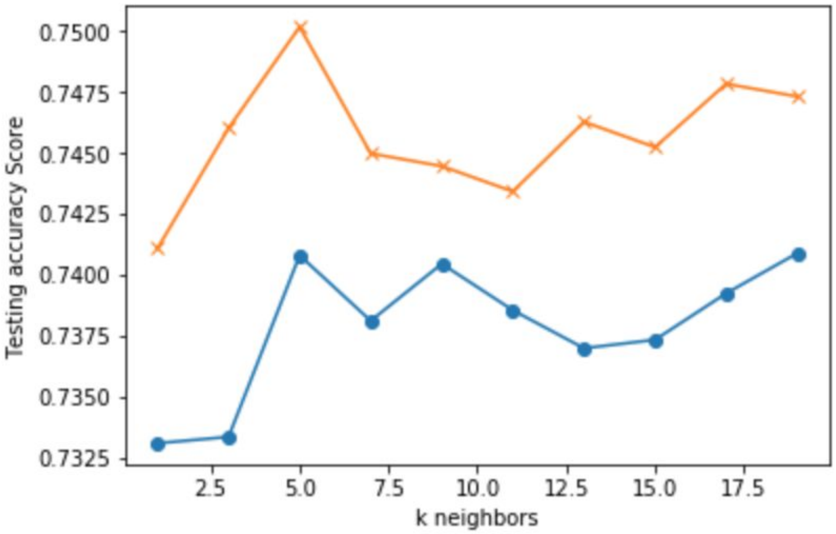
K-nearest neighbor

k: 7, Train/Test Score: 0.866/0.838, standar scaler



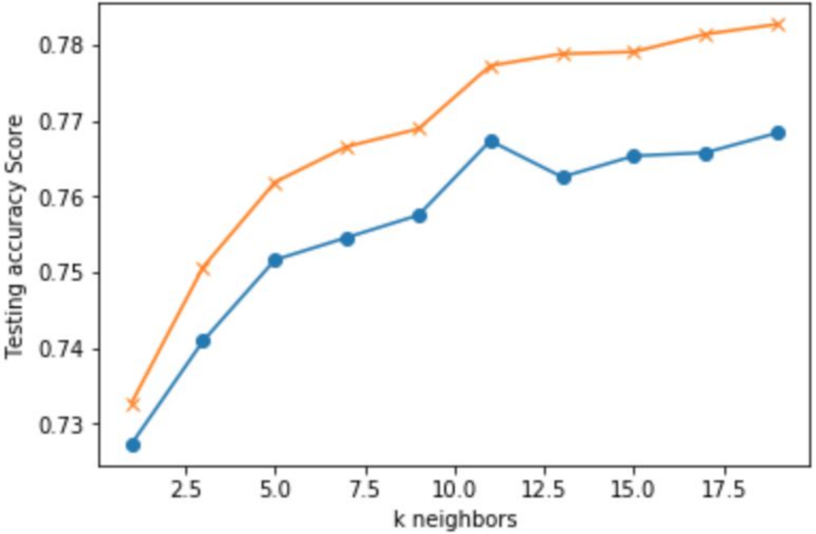
K-nearest neighbor

k: 5, Train/Test Score: 0.741/0.750, Robust Scaler



K-nearest neighbor

k: 11, Train/Test Score: 0.767/0.777, Quantile tranformer.



SVM model

	precision	recall	f1-score	support
1	0.79	1.00	0.88	3021
0	0.00	0.00	0.00	822
micro avg	0.79	0.79	0.79	3843
macro avg	0.39	0.50	0.44	3843
weighted avg	0.62	0.79	0.69	3843

SVM Naive Bayes Model

- Accuracy : 0.7816809784022899

Regression

- Training Data Score: 0.7690839694656488
- Testing Data Score: 0.7868852459016393

Neural Networks

- Initial training score

  loss : ~0.47, accuracy : ~0.75

- Final score :
- 3843/3843 - 0s - loss: 0.3627 - categorical_accuracy: 0.8332
- Normal Neural Network - Loss: 0.3626526763935089, Accuracy: 0.8332032561302185

How to optimize accuracy and reduce loss

- Change Epoch number : that didn’t really help
- Add layers and units : that helps a little
- Activation parameters : find the best combination, helps a lot
- Add optimizers : adam, RMSPropOptimizer,GradientDescentOptimizer
- Reclean data: remove outliers , helps the most
- Retrain model

Reference

- Özdemir, Ahmet Turan, and Billur Barshan. "Detecting Falls with Wearable Sensors Using Machine Learning Techniques." Sensors (Basel, Switzerland) 14.6 (2014): 10691–10708. PMC. Web. 23 Apr. 2017.
- https://www.kaggle.com/pitasr/falldata
- Centers for Disease Control and Prevention
- National Institute of Senior Center
- https://www.ncoa.org/news/resources-for-reporters/get-the-facts/falls-prevention-facts/