Human Activity Recognition Using Machine Learning

Jay Upadhyay (jxu9414), Sahana Murthy (sam2738), Sanchitha Seshadri (ss9886)

Problem Definition:

- Smart-phones are now ubiquitous
- Collect vast amount of reliable data
- Contain an accelerometer - tracks spatial orientation
- Can be used to infer physical activity
- Sitting, standing, walking, sit-to-stand, etc.

Motivation:

- Technology gives us unprecedented access to data
- Data to track lifestyle changes for medical and simply wellness purposes

The data:

- HAPT Dataset, obtained from the UCI Machine Learning Repository webpage. The dataset contains
- 10929 instances of real-world data
- Each instance has 561 attributes.
- http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Post ural+Transitions

Intuition:

- Several activities the person may be performing - classification problem.
- Need to infer activity using features / attributes in the data.
- Choose appropriate algorithm to solve this
- examine three algorithms and analyze which works best for given problem

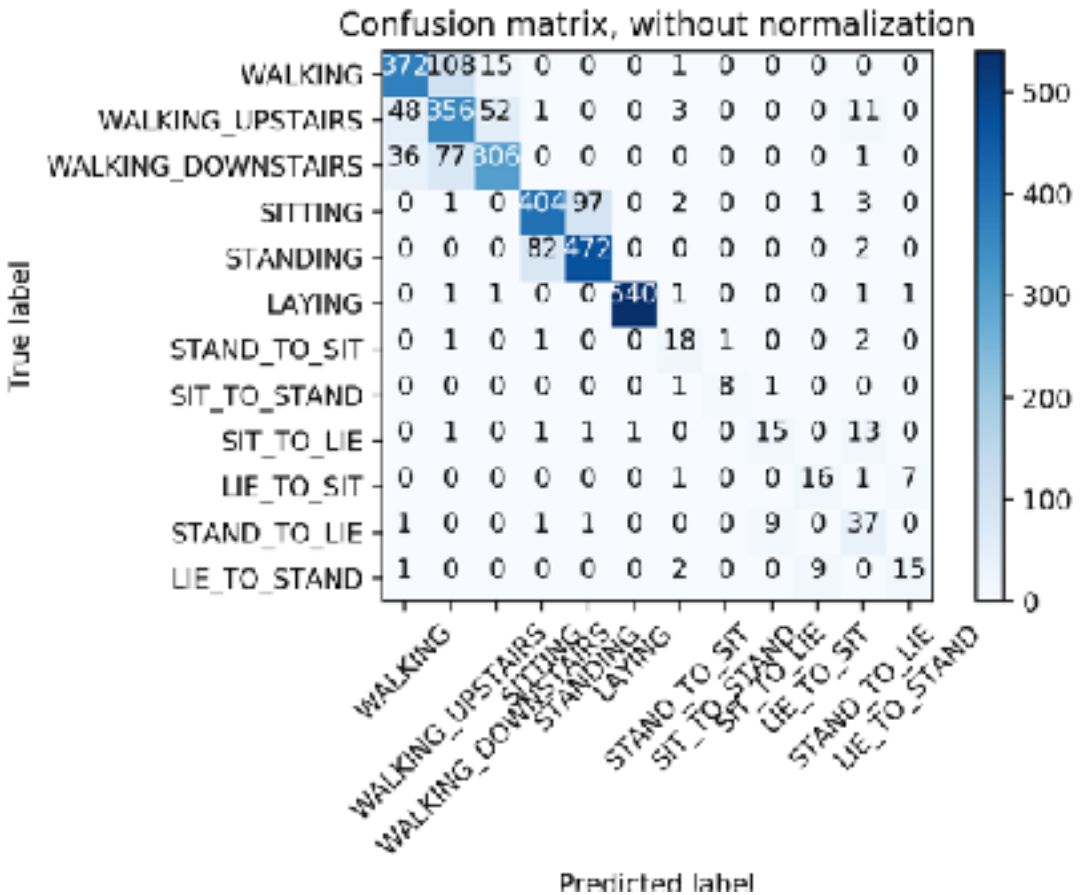
Decision Tree Algorithm

- The dataset has 561 attributes - need to be used to make a decision
- Information gain to quantify amount of learning at each stage
- Calculates IG for every attribute - chooses best feature at each step
- Helps select the features useful / most relevant for problem

Experimental Results and Analysis:

- Number of correctly classified records: 2562
- Accuracy: 81.025%
- Mean Squared Error: 0.577

The confusion matrix obtained for this model:



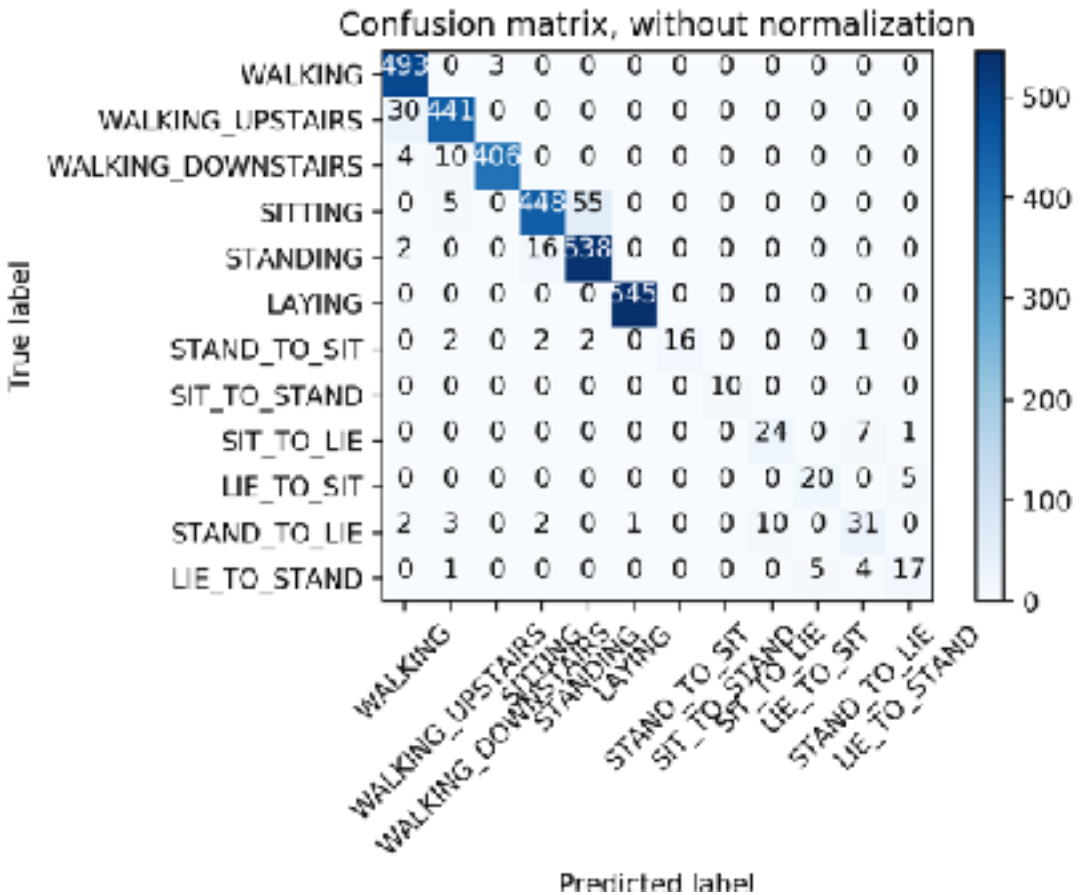
Logistic Regression

- Useful when the data is categorical in nature.
- Finds line to separate data to classify different instances correctly - logistic function
- Estimates probabilities of instance pertaining to each activity
- Predicted event - one with largest probability calculated.

Experimental Results and Analysis:

- Number of correctly classified records: 2989
- Accuracy: 94.53%
- Mean Squared Error: 0.338

The confusion matrix obtained for this model:



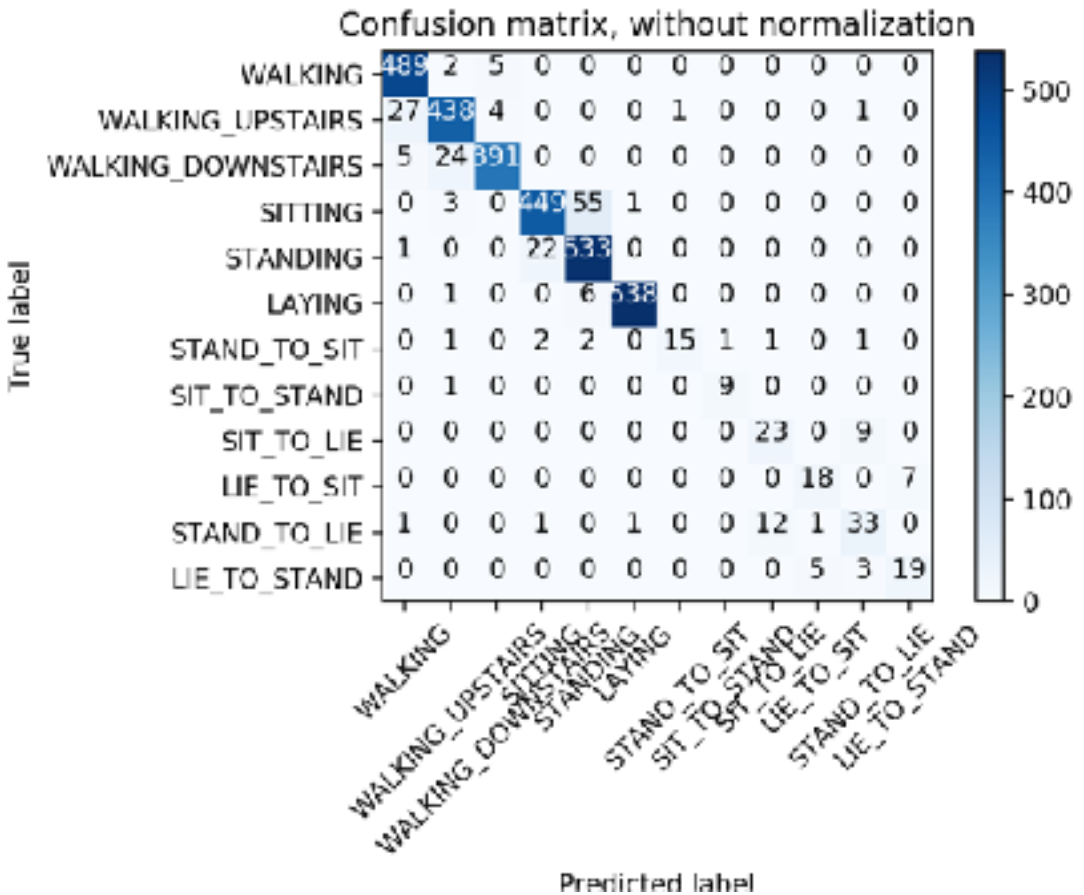
Multi Layered Perceptron

- 561 attributes - not sure which ones actually useful
- Neural network layers work to determine most useful ones
- Optimizes the log-loss function for classification
- Uses the stochastic gradient descent.

Experimental Results and Analysis:

- Number of correctly classified records: 2914
- Accuracy: 92.16%
- Mean Squared Error: 0.285

The confusion matrix obtained for this model:



- Started with Decision Tree - mediocre accuracy
- Then tried Multi-Layer Perceptron Classifier - recorded boost in accuracy by ~10%
- Implemented Logistic regression - most accurate model for problem (accuracy ~94%) but
- Mean Squared error rose!
- Logistic regression took longest time

Conclusion:

- MLP Classifier or Logistic Regression best for this problem
- Identified a time-accuracy trade off
- MLP faster although slightly less accurate
- Logistic Regression slower but slightly more accurate
- Mean Squared Error may not be a good parameter to determine the effectiveness of model

References

- http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions

- Lecture slides

Thank You / Questions?