UNIVERSITY of WASHINGTON

NBA STATISTIC ANALYSIS

KuanHsun Lu, YiCheng Chen

Task Description

- Task1: All star prediction
- Task2: Players points per game prediction
- Task3: Playing styles vs actual position

Approach

- Task1:
  - Use selected and normalized features to train model, then fit test data to see if these players are qualified or not.
  - Player efficiency rating (PER) higher than 20 will be selected as All star player.
  - Fit different classifiers to find out which gives the best performance.

- Task2: Players per game prediction
  - Use selected and normalized features to train model.
  - Use two types of normalization standardize and MinMax scalar.
  - Models are built by linear regression and neural network
  - Neural Network Structure: 4 layers, batch size: 10, epoch: 200

- Task3: Playing styles versus actual position
  - Use K means cluster to find out the similarity of each player
  - Choose K = 5 (number of positions)

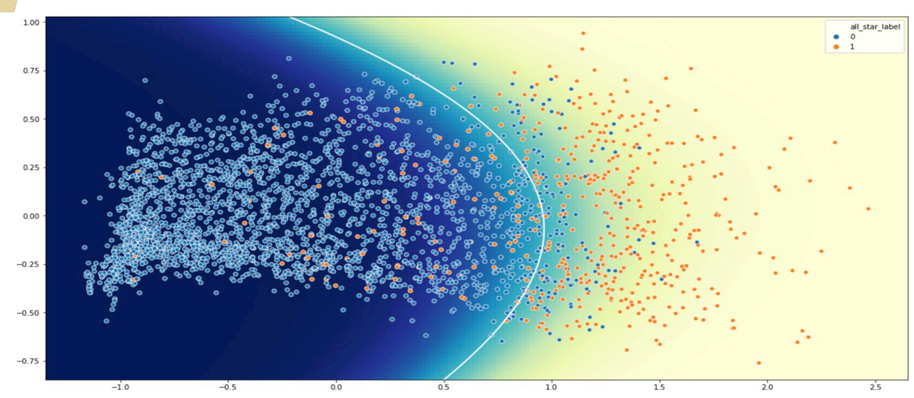
Experimental Paradigm

- Data
  - Train data: 1980~2000 players stats
  - Test data: 2013~2017 players stats
  - Cluster data: 2017 players stats

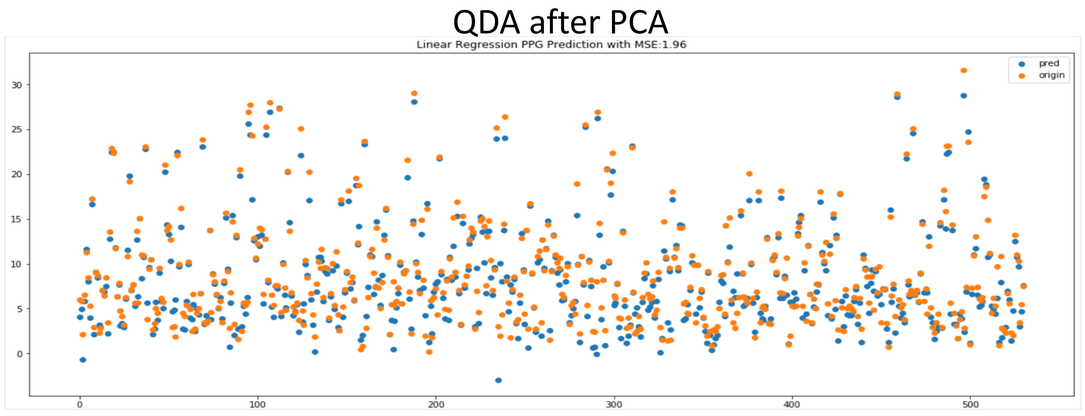
- Software
  - Python 3.6
  - Package: keras, sklearn

	KNN	Gaussian NB	QDA	QDA With PCA	LDA	SVC	Decision Tree
Number	434	744	2978	499	419	427	455
Accuracy	94%	86.19%	14%	92%	95%	96%	97.31%
Log Loss	0.67	2.8	29.4	0.23	0.1	0.07	0.92
Precision	94%	91%	2%	88%	96%	97%	97%
Recall	94%	86%	15%	68%	96%	97%	97%

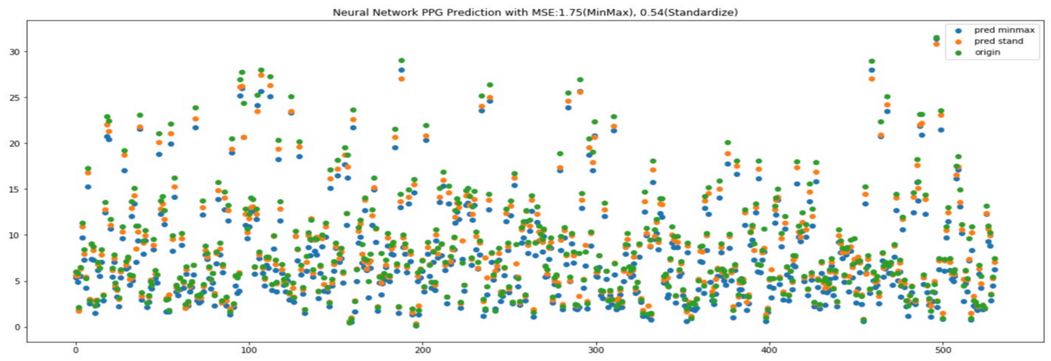
Experimental Results



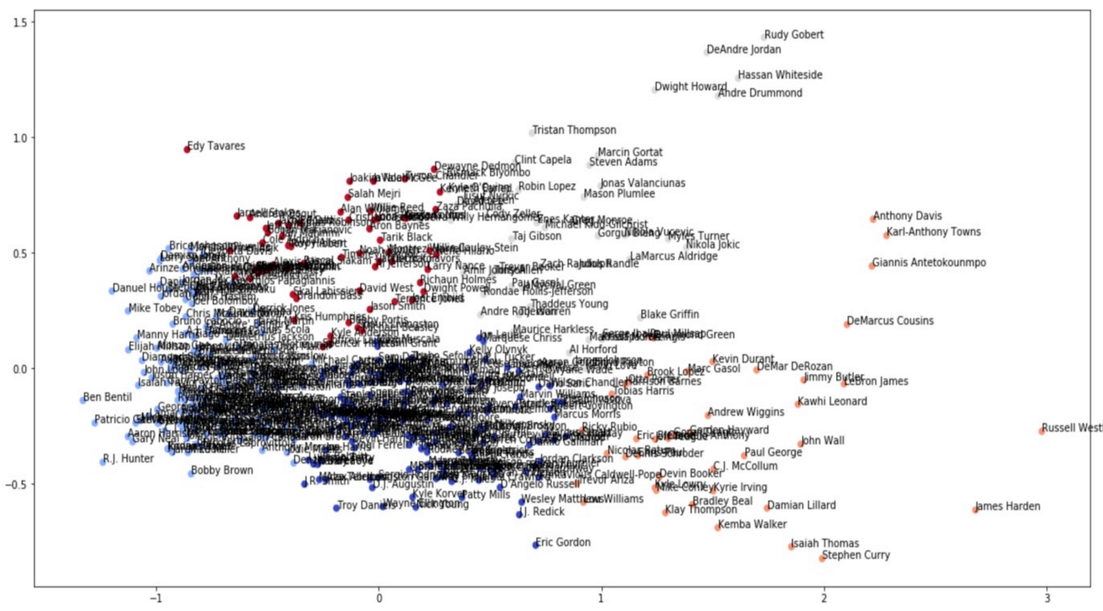
QDA after PCA



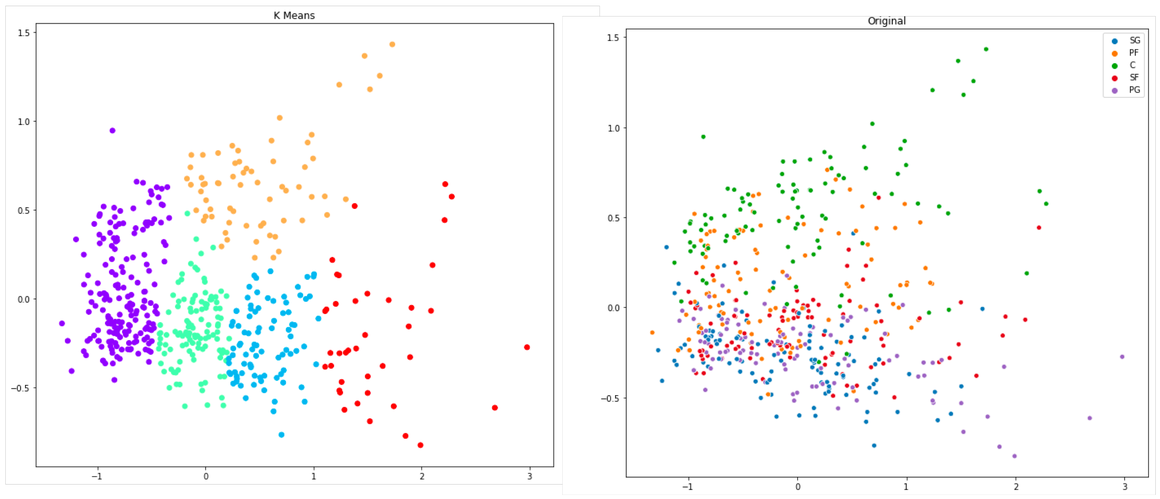
Linear Regression MSE: 1.96



NN MSE: 0.58(Standardize), 1.98(MinMax scalar)



Players name plot by clusters



K Means

Players position plot

Summary of Findings

- The majority class will dominant in SVC
- PCA before QDA
- Standardization outperforms MinMax scalar in neural network
- Neural network gives better performance than linear regression
- Interesting observations in NBA