Date_A_Scientist

Machine Learning Fundamentals

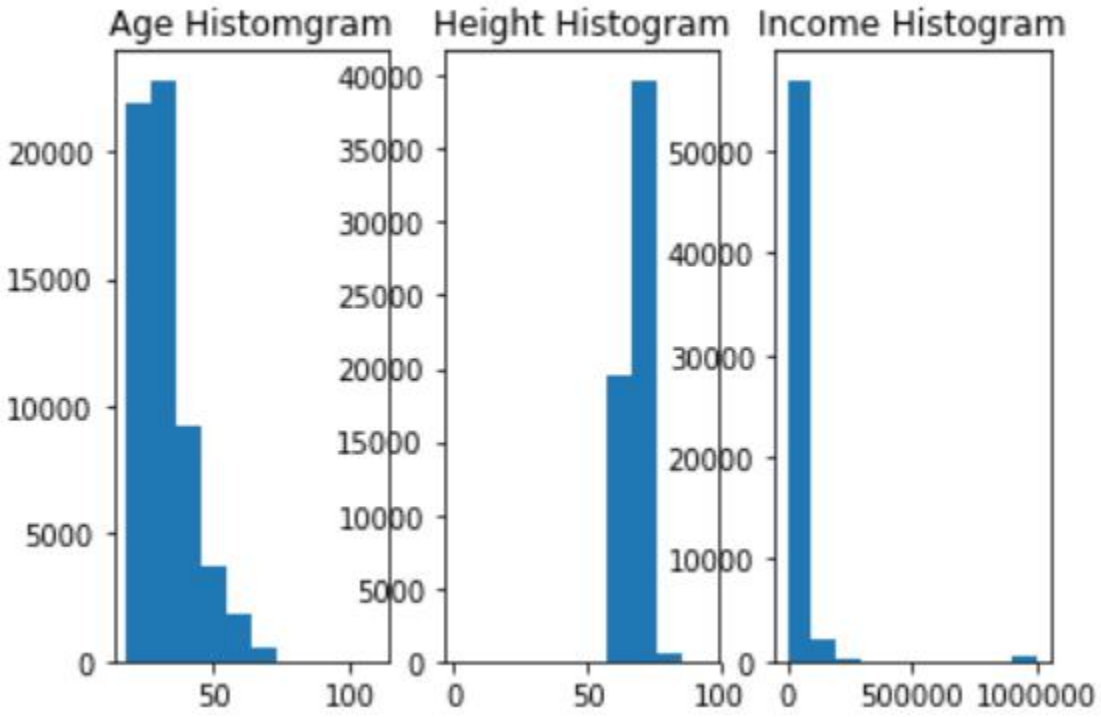
Ravi Siddamsetty

11/14/2018

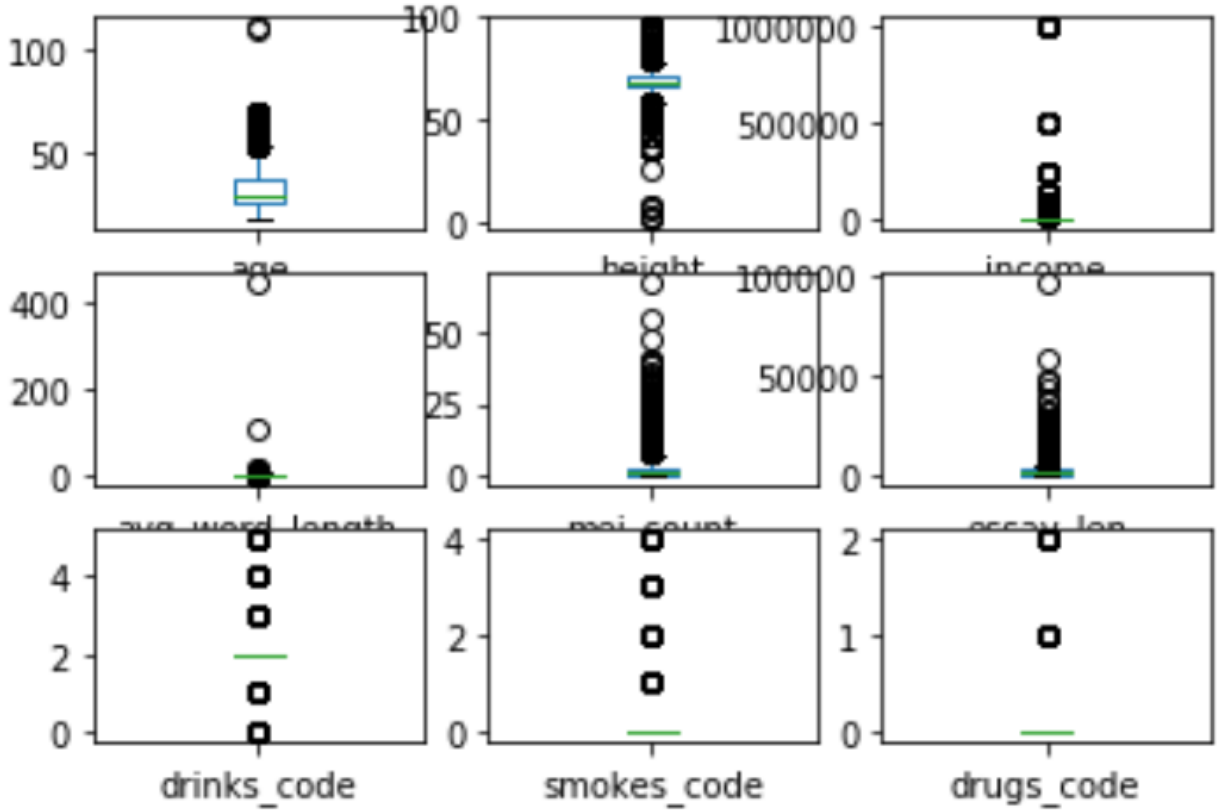
Numeric Data Exploration df.describe()

	age	height	income
count	59946.000000	59943.000000	59946.000000
mean	32.340290	68.295281	20033.222534
std	9.452779	3.994803	97346.192104
min	18.000000	1.000000	-1.000000
25%	26.000000	66.000000	-1.000000
50%	30.000000	68.000000	-1.000000
75%	37.000000	71.000000	-1.000000
max	110.000000	95.000000	1000000.000000

Data Exploration Histograms



Data Exploration



Questions

- Classify Zodiac Sign using diet, drinks, smokes, drugs
- Classify Diet using drinks, smokes, drugs
- Predict age with the frequency of "I" or "me" in essays?

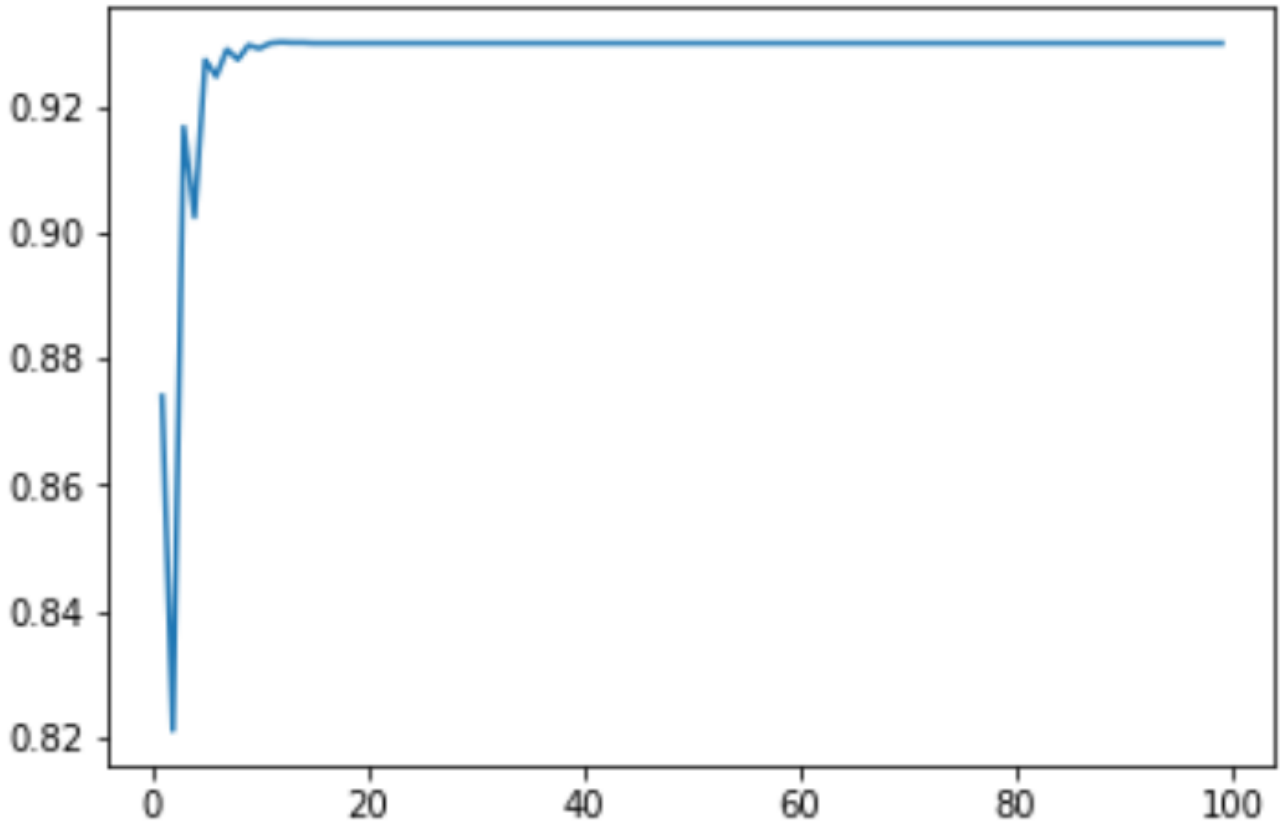
Compare classification models: classify zodiac sign

```
LR: 0.249854 (0.006219)

C:\Users\siddam\AppData\Local\Cont
ors do not sum to 1. Renormalizing
  UserWarning)

LDA: 0.249562 (0.006389)
KNN: 0.181812 (0.005921)
CART: 0.128034 (0.003928)
NB: 0.226228 (0.008221)
```

Compare classification models: classify diet code

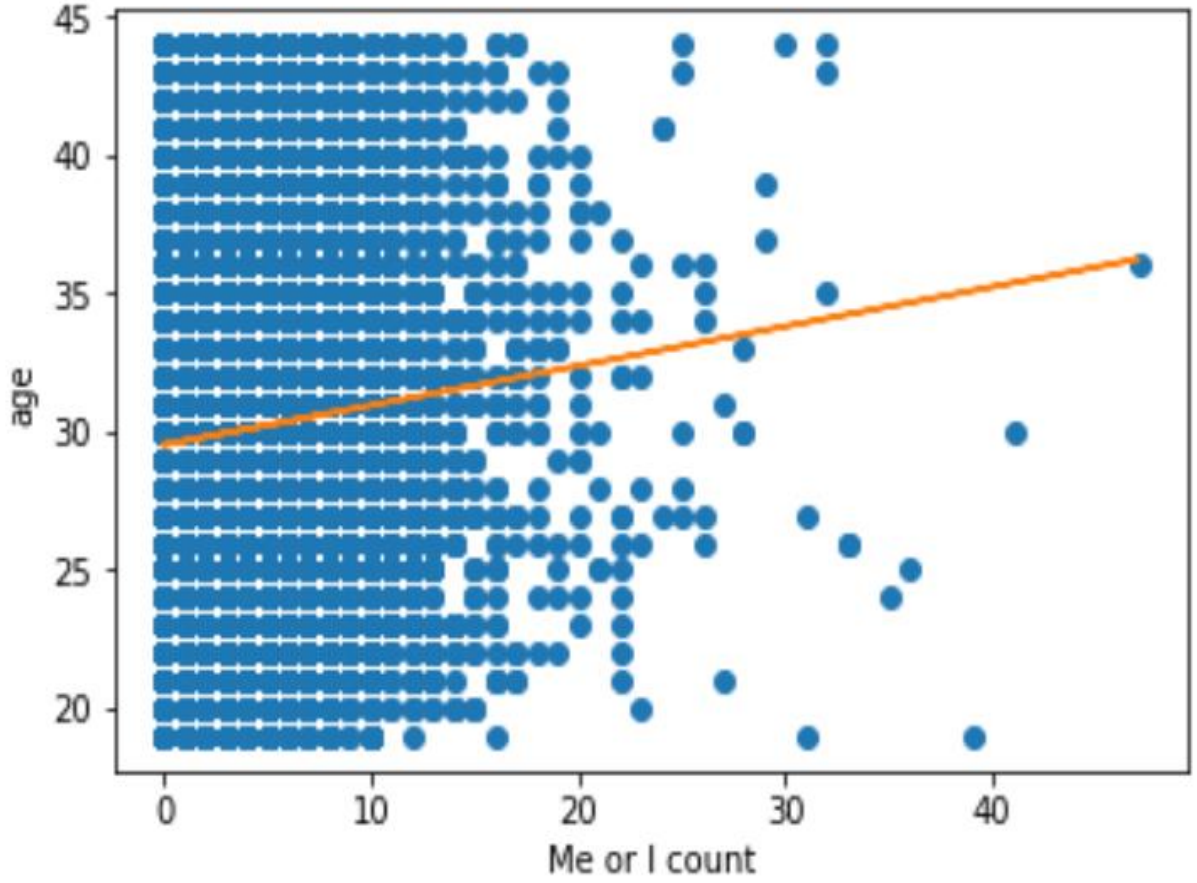
```
LR: 0.931062 (0.003581)
KNN: 0.927663 (0.003634)
CART: 0.871716 (0.003844)
NB: 0.914338 (0.004575)
SVM: 0.931062 (0.003581)
```

KNN Classifier for Diet Code



Results: Regression

- X = Me or I count
- Y = Age
- Slope [[0.14292244]]
- Intercept [29.52853805]



Results: KNN Classifier for Zodiac sign code

- The accuracy we would expect from predicting a Zodiac sign by randomly selecting one would be 1/12, or 0.0833. Our model did not significantly outperform this number. We were unimpressed.

```
LR: 0.249854 (0.006219)
KNN: 0.181812 (0.005921)
CART: 0.127638 (0.004209)
NB: 0.226228 (0.008221)
```

Results KNN Classifier for Diet Code

- The accuracy for diet code is ½ by random and our model accuracy is 93%