CS677A3Fall2023WillMohrHomework6

Homework 6

Logistic Regression

1) SPY: logit(p) = 0.10 + 3.07*mean_return -.28*volatility GME: logit(p) = .39 + 2.24*m_r + .34 * vol

2)Accuracy GME: 82.69% SPY: 96.15%

3) a) SPY: [[19, 2], [0, 31]] b) GME: [[26, 8],[1, 17]]

4) Y2 TPR: {spy: 1.0, gme: 0.94} TNR: {SPY: 0.9, gme: .76}

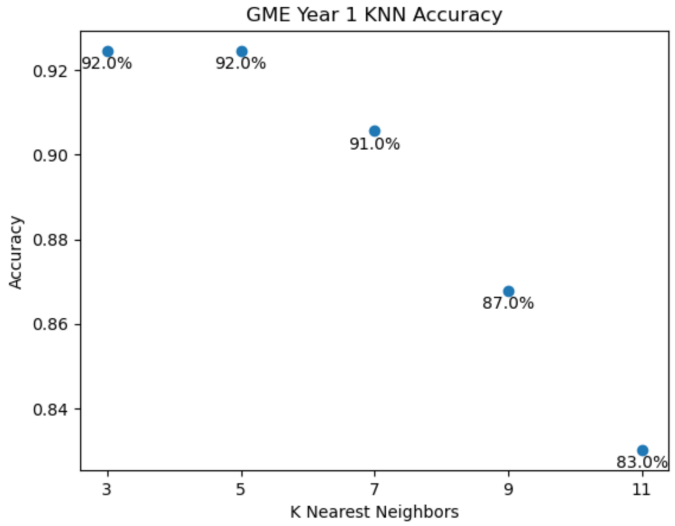
5) spy: logreg "green" prediction strategy: $165.06, b&h: $129.0

Gme: logreg: $44,369.46 b&h: $824

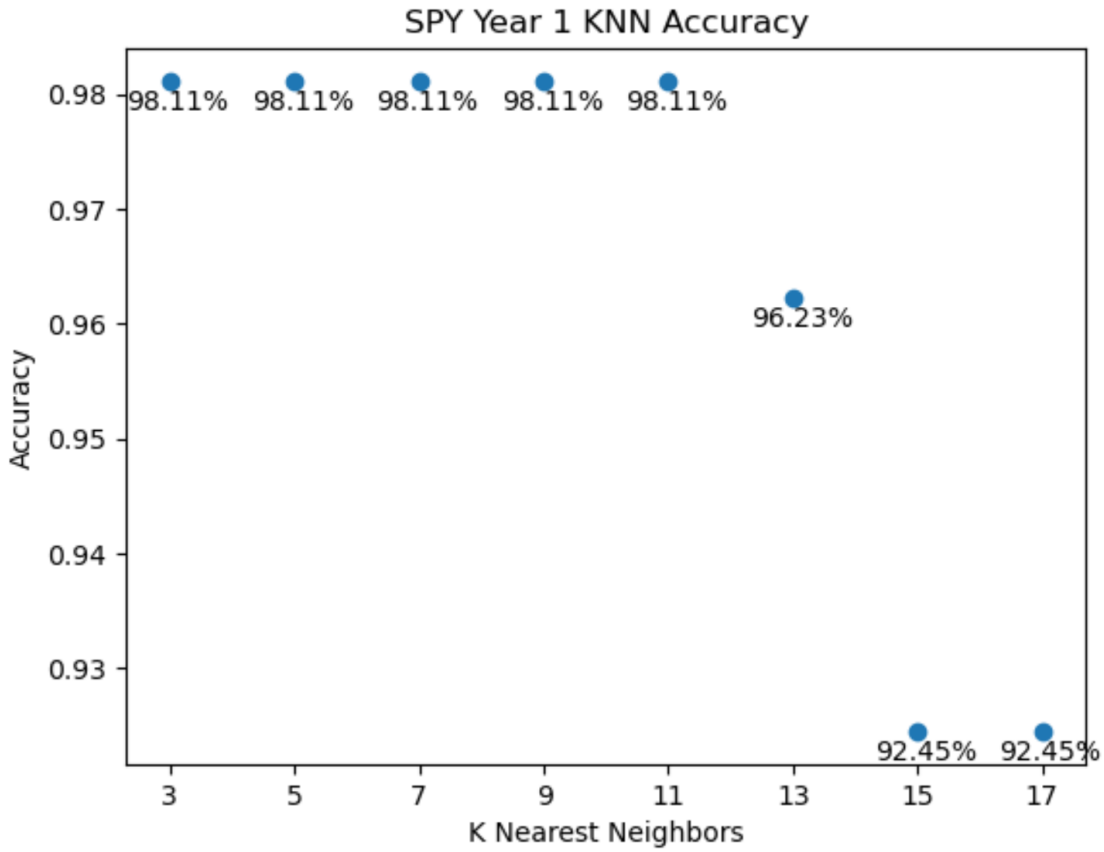
KNN

1)

GME: max accuracy is 92% at K = 3, 5. SPY Max accuracy: 98.11% at k = 3… 11



SPY:



2) GME: 86.5%, SPY: 98.1%
3) GME: [[17, 1], [6, 28]] SPY:
4) GME: True positive rate is 17/18, and the true negative rate is 28/34 SPY:
True positive rate is 20/21, and the true negative rate is 31/31.

5) GME: $100 invested in gme at the beginning of 2021 would result in $45,809.4 by the end of 2021 if traded according to "green" predictions. Admittedly, knowing return and volatility in advance is not possible in the real world. The buy-and-hold strategy would have resulted in $824.0

GME:

$100 invested in spy at the beginning of 2021 would result in $163.66 by the end of 2021 if traded according to "green" predictions. Admittedly, knowing return and volatility in advance is not possible in the real world. The buy-and-hold strategy would have resulted in $129.0

Bernford's Law:

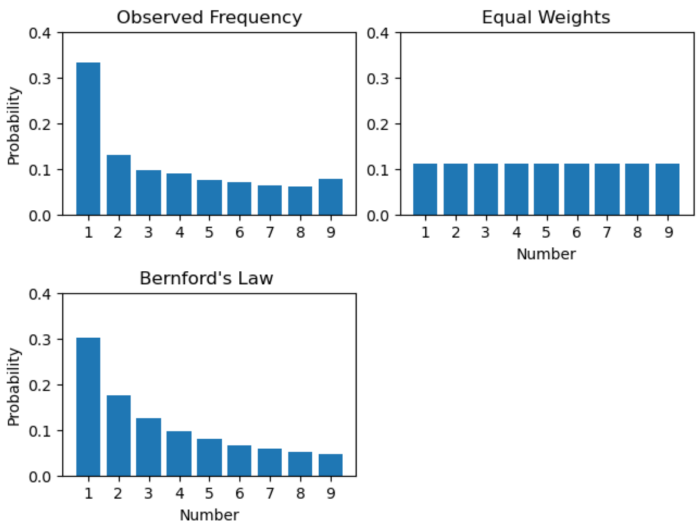
2) 7 categories

3)
Category with least subcategories:
New Arrivals
Category with most subcategories:
Clothing
4)

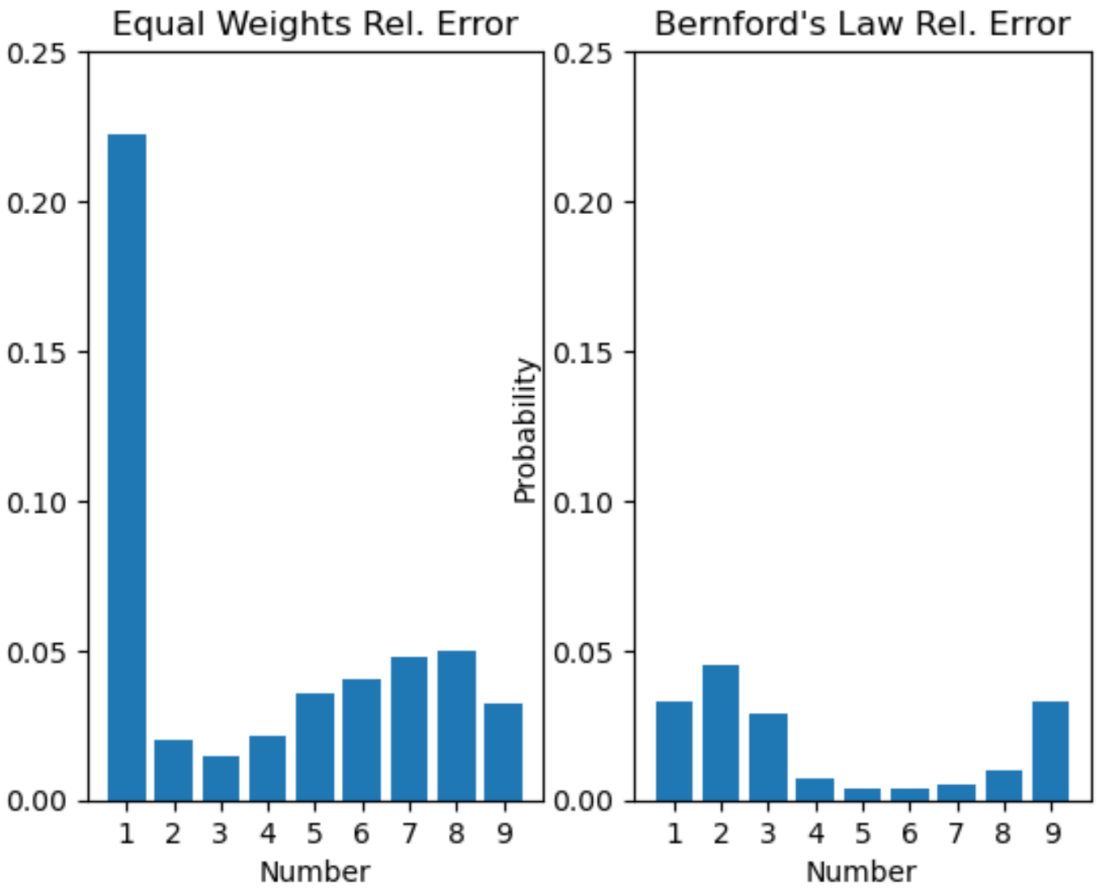
:

CATEGORY	median	mean
Accessories	198.0	661.102010
Clothing	89.0	215.042551
Handbags	698.0	1017.721767
Jewelry	325.0	1003.327448
Maternity	68.0	82.381425
New Arrivals	89.0	258.012151
Shoes	130.0	185.105339

5)



6)



7)
Bernstein's model (RMSE of .024) is closer to the real distribution than the equal weighting model (RMSE of .081)

8)

{'Clothing': {'1': 0.2182831026761132,
  '2': 0.1290609742667303,
  '3': 0.12919863441136523,
  '4': 0.1132759443485922,
  '5': 0.10714547924085019,
  '6': 0.1012260930215484,
  '7': 0.07813589809478359,
  '8': 0.05980874417238721,
  '9': 0.06386512976762967},
 'Jewelry': {'1': 0.2514291250847786,
  '2': 0.15457158544068728,
  '3': 0.12149985466524561,
  '4': 0.14840293253237735,
  '5': 0.0924329037883926,
  '6': 0.07118173303620451,
  '7': 0.056486774537351035,
  '8': 0.04779898588638052,
  '9': 0.0561961050285825},
 'Maternity': {'1': 0.13778950454412195,
  '2': 0.03928466725300499,
  '3': 0.056874816769275874,
  '4': 0.20433890354734682,
  '5': 0.18059220170038112,
  '6': 0.10290237467018469,
  '7': 0.09586631486367635,
  '8': 0.08384637936089123,
  '9': 0.09850483729111698}}

Bernfords = [0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046],
P = [0.111, 0.111, 0.111, 0.111, 0.111, 0.111, 0.111, 0.111, 0.111]

Clothing rmse: 0.045
Jewelry rmse: 0.062
Maternity rmse: 0.051

From above we can see that clothing has an rmse that is closest to equal weighting.