**DMG Assignment**

**By – Vineet Kapoor - 703222761**

P1 : IRIS –HIERARCHICALFISHER

CreatethesecondFisherprojectionbytryingtodiscriminateclass1from class 2 (the original two similarclasses).

P2 : MUSHROOM informationgain

Take the MUSHROOM training data.Thereare20+features and 2 classes. We want to find the BEST feature using the three purity measures: Accuracy, GiniIndex and Entropy.

Weighted Entropy - best attribute is V6

'V6': 0.093676408087834

Gini Index - best attribute is V6

'V6': 0.028742240686071896

Accuracy – best attribute is V10

'V10': 0.788057876502955

Measure the Informationgainduetoeachfeature.Generate atablewiththefollowingcolumns:

¤Feature\_name

¤Accuracy

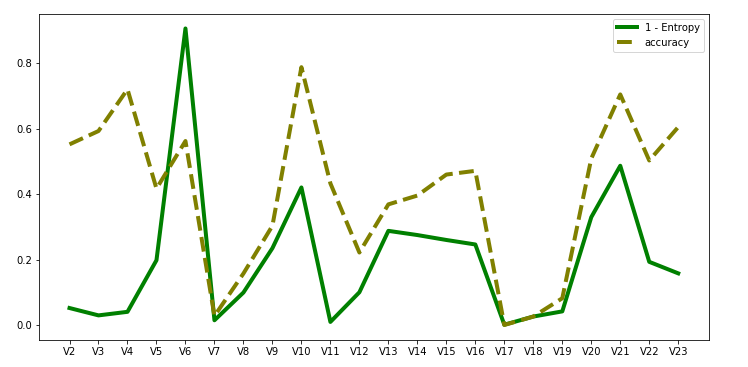
¤GINIindex

¤1-Entropy(NOTE:Uselog\_kforafeaturewithkvalues)

|  | **information gain** | **accuracy** | **Gini index** | **feature** | **1 - Entropy** |
| --- | --- | --- | --- | --- | --- |
| **0** | 0.051022 | 0.552068 | 0.468190 | V2 | 0.051818 |
| **1** | 0.028518 | 0.592827 | 0.480097 | V3 | 0.029314 |
| **2** | 0.039399 | 0.720399 | 0.473230 | V4 | 0.040195 |
| **3** | 0.197132 | 0.416752 | 0.370817 | V5 | 0.197928 |
| **4** | 0.905527 | 0.562258 | 0.028742 | V6 | 0.906324 |
| **5** | 0.013589 | 0.027512 | 0.491292 | V7 | 0.014385 |
| **6** | 0.098365 | 0.157326 | 0.440331 | V8 | 0.099161 |
| **7** | 0.233949 | 0.303240 | 0.352344 | V9 | 0.234745 |
| **8** | 0.419571 | 0.788058 | 0.266574 | V10 | 0.420367 |
| **9** | 0.008563 | 0.432851 | 0.493528 | V11 | 0.009359 |
| **10** | 0.099184 | 0.221520 | 0.439685 | V12 | 0.099980 |
| **11** | 0.286869 | 0.368657 | 0.324875 | V13 | 0.287666 |
| **12** | 0.274080 | 0.395557 | 0.332647 | V14 | 0.274877 |
| **13** | 0.259021 | 0.459751 | 0.359564 | V15 | 0.259817 |
| **14** | 0.245242 | 0.471164 | 0.365346 | V16 | 0.246038 |
| **15** | 0.000000 | 0.000000 | 0.499448 | V17 | 0.000796 |
| **16** | 0.024949 | 0.025678 | 0.487114 | V18 | 0.025746 |
| **17** | 0.040745 | 0.081720 | 0.475069 | V19 | 0.041541 |
| **18** | 0.328666 | 0.508661 | 0.311420 | V20 | 0.329462 |
| **19** | 0.485882 | 0.704708 | 0.214691 | V21 | 0.486679 |
| **20** | 0.192191 | 0.502751 | 0.387116 | V22 | 0.192987 |
| **21** | 0.156956 | 0.605054 | 0.402403 | V23 | 0.157753 |

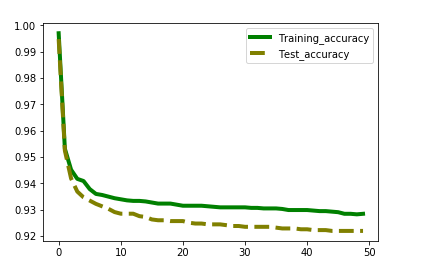
Plotaccuracyvs.1–Entropyscatterplotwhereeachpointis

afeature.



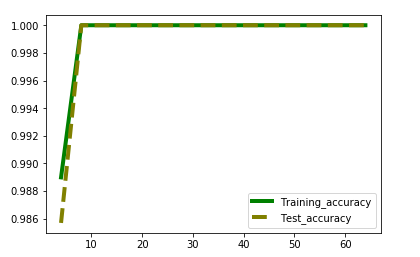
InNaïveBayesclassifierplotthevalueoflambda(x-axis)for Laplaciansmoothingagainsttrainingandtestsetaccuracy.

¤ Lambda = 0, 1, 2, …,50



FordecisiontreeclassifierplottheSizeThreshold(x-axis) againsttrainingandtestsetaccuracy.

¤ SizeThreshold = 4, 8, 12, 16, 20, …,64.



P4 : MNISTBayesian

Compare the test accuracies of the four classifiers and comment.

Accuracy for Diagonal Covariance Matrix on D1 data

0.09892517569243489

Accuracy for Full Covariance Matrix on D2 data

0.14452252997106244

Accuracy for Diagonal Covariance Matrix on D1 data

0.1090533278214138

Accuracy for Full Covariance Matrix on D2 data

0.10892930963207938

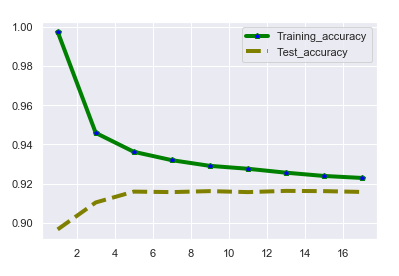
P5 : MNIST –kNN /Parzenwindow

¤Plot training and test accuracy with these values of k on x axis

On D1 dataset

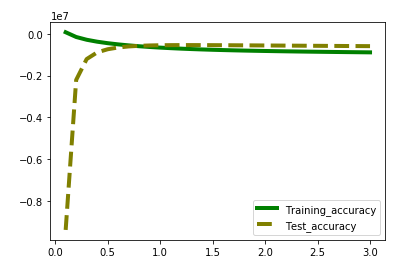


On D2 dataset



Plottrainingandtestaccuracieswiththesevaluesofsigma.

On D1 dataset



On D2 dataset

Commenton

theoptimalkandoptimalsigmaandcompare those classifiers across D1 and D2 and see which one has highest testaccuracy.