Task 1 Write Up

Group 443

2024-12-08

Task 1: Binary Classification

Abstract Task 1

- abstract/overview of task

Task 1.1 Data Description
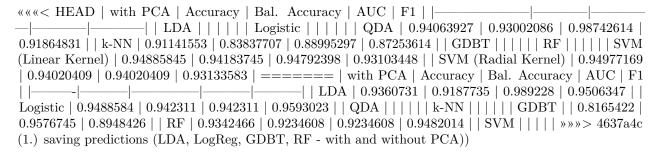
- summary statistics
- plots

Task 1.2 Summary Table without PCA

- short intro (1-2 sentences what every classifier does)
- after showing summary table, add explanation to specific occurances like the QDA or kNN or SVM

w/o PCA	Accuracy	Bal. Accuracy	AUC	F1
LDA	0.9415525	0.9406989	0.9407	0.9525926
Logistic	0.6392694	0.6413256	0.6596597	0.6857597
QDA				
k-NN				
GDBT	0.906821	0.8856173	0.9723815	0.9283372
RF	0.9351598	0.9170923	0.9170923	0.9500352
SVM				

Task 1.2 Summary Table with PCA

- explanation of what PCA does and how it is essentially applied for all classifiers

«««< HEAD | with PCA | Accuracy | Bal. Accuracy | AUC | F1 | |————————|————|————————|————————|————————| | LDA | | | | | | Logistic | | | | | | QDA | 0.94063927 | 0.93002086 | 0.98742614 | 0.91864831 | | k-NN | 0.91141553 | 0.83837707 | 0.88995297 | 0.87253614 | | GDBT | | | | | | RF | | | | | | SVM (Linear Kernel) | 0.94885845 | 0.94183745 | 0.94792398 | 0.93103448 | | SVM (Radial Kernel) | 0.94977169 | 0.94020409 | 0.94020409 | 0.93133583 | ======= | with PCA | Accuracy | Bal. Accuracy | AUC | F1 | |————-|————|————————|————————|————————| | LDA | 0.9360731 | 0.9187735 | 0.989228 | 0.9506347 | | Logistic | 0.9488584 | 0.942311 | 0.942311 | 0.9593023 | | QDA | | | | | | k-NN | | | | | | GDBT | | 0.8165422 | 0.9576745 | 0.8948426 | | RF | 0.9342466 | 0.9234608 | 0.9234608 | 0.9482014 | | SVM | | | | | »»»> 4637a4c (1.) saving predictions (LDA, LogReg, GDBT, RF - with and without PCA))

Task 1.3 Improved Classifiers

- Summary Table: of F1 before and F1 after improvement

	F1 Score before	F1 Score after
QDA/LDA??		
Logistic (with Ridge)		
RF (with AdaBoost)		

- explanation how each classifier was improved and why this improvement method was chosen

Task 1.4 Final predict function

- summary and discussion of the results

	Accuracy	Bal. Accuracy	AUC	F1 Score
mypredict()				

- present in one row all performance metrics of chosen prediction function