Assignment 3

Alex Radecki

Mar 22^{nd} 2023

1 Bagging Classifier Experiment Results

Normalization	Bagging Classifier(BG) parameter	Precision	Recall	Accuracy	F2Score %
No preprocessing	BG with estimator 5	0.875	1.0	0.9667	0.9722
Minmax	BG with estimator 5	0.8182	0.9643	0.9417	0.931
Z-score	BG with estimator 5	0.9032	1.0	0.975	0.979
No preprocessing	BG with estimator 10	1.0	0.9643	0.9917	0.9712
Minmax	BG with estimator 10	0.9643	0.9643	0.9833	0.9643
Z-score	BG with estimator 10	0.875	1.0	0.9667	0.9929
No preprocessing	BG with estimator 25	0.875	1.0	0.9667	0.9722
Minmax	BG with estimator 25	0.8235	1.0	0.95	0.9589
Z-score	BG with estimator 25	0.9655	1.0	0.9917	0.9929
No preprocessing	BG with estimator 50	0.9333	1.0	0.9833	0.9859
Minmax	BG with estimator 50	0.9333	1.0	0.9833	0.9859
Z-score	BG with estimator 50	0.9655	1.0	0.9917	0.9929
No preprocessing	BG with estimator 100	0.9333	1.0	0.9833	0.9859
Minmax	BG with estimator 100	1.0	1.0	1.0	1.0
Z-score	BG with estimator 100	0.9655	1.0	0.9917	0.9929

2 Random Forest Experiment Results

Normalization	Estimators	Max depth	Precision	Recall	Accuracy	F2Score %
No preprocessing	3	3	0.9	0.87	0.9417	0.8847
No preprocessing	3	50	0.9083	0.8656	0.9417	0.8864
No preprocessing	50	3	0.91	0.89	0.95	0.9
No preprocessing	100	100	0.9667	0.86	0.9667	0.9102
Minmax	3	3	0.8981	0.95	0.9583	0.9233
Minmax	3	50	0.9167	0.8767	0.95	0.8962
Minmax	50	3	0.8667	0.9095	0.9583	0.8876
Minmax	100	100	0.9314	0.885	0.95	0.9076
Z-score	3	3	0.8167	0.83	0.9333	0.8233
Z-score	3	50	0.8629	0.8248	0.925	0.8434
Z-score	50	3	0.975	0.8167	0.95	0.8888
Z-score	100	100	0.96	0.8362	0.95	0.8938

3 Discussion

For the bagging classifier experiment, estimators of 50 and 100 seemed to perform the best, which is what led me to use them as values for my parameters (n_estimators and max_depth) for the Random Forest experiment. In both, it wasn't immediately obvious which preprocessing method performed the best, but overall it would appear that z-score overall performed the best, which is in line with expectations. While increasing the number of estimators appeared to improve the performance of the bagging classifier, it was difficult to notice any trends for the Random Forest experiment. Some less extreme parameters may have helped better visualize any trends, as my thought process was an extreme parameter would make noticing any changes easier, which wasn't the case. Increasing the number of estimators helps mitigate overfitting at the cost of complexity, while increasing the depth leads to more variance and overfitting. In other words, my choice of 100 estimators and max depth of 100 was likely pointless, in hindsight. Also, having a large number for estimators and max depth may have caused a drop in model performance, preventing me from seeing any trends I would have otherwise noticed had I chosen smaller parameters and experimented with different combinations.