DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Homework Assignment No. 08:

HW No. 08: LDA, K-NEAREST NEIGHBORS AND K-MEANS CLUSTERING

submitted to

Professor Joseph Picone
ECE 8527: Introduction to Pattern Recognition and Machine Learning
Temple University
College of Engineering
1947 North 12th Street
Philadelphia, Pennsylvania 19122

April 14, 2024

prepared by:

Leo Berman
Email: leo.berman@temple.edu

A. DATA TABLE

DS	System	Training Data	Train	Dev	Eval
08	KNN(PAPER)	/train	76.52	73.38	35.82
	RNF(PAPER)		70.77	60.55	40.23
	CI-PCA		36.07	36.02	31.76
	QDA		49.92	49.78	46.34
	LDA		36.09	36.01	31.94
	KNN		75.19	74.14	35.92
	RNF		100.00	71.18	36.86
	KNM		35.47	35.42	33.22
	CI-PICA	/train + /dev	36.08	36.01	31.76
	QDA		49.93	49.78	46.34
	LDA		36.09	36.02	31.96
	KNN		83.81	84.26	38.10
	RNF		100.00	100.00	36.76
	KNM		33.98	33.94	36.67
09	KNN(PAPER)	/train	97.89	96.19	83.37
	RNF(PAPER)		97.94	96.18	81.68
	CI-PCA		96.20	91.82	97.36
	QDA		96.87	93.04	97.13
	LDA		96.20	91.82	97.36
	KNN		98.13	96.31	83.24
	RNF		100.00	95.92	81.90
	KNM		87.13	84.69	90.96
	CI-PICA	/train + /dev	95.93	91.80	97.11
	QDA		97.15	93.84	96.68
	LDA		95.93	91.80	97.11
	KNN		97.86	98.62	83.60
	RNF		100.00	100.00	80.89
	KNM		84.47	84.12	89.51

DS	System	Training Data	Train	Dev	Eval
10	KNN(PAPER)	/train	92.37	61.17	66.56
	RNF(PAPER)		97.85	20.27	66.72
	CI-PCA		53.02	51.13	48.90
	QDA		83.92	55.24	65.89
	LDA		53.02	50.13	48.90
	KNN		92.48	61.37	66.53
	RNF		100.00	60.21	66.68
	KNM		50.00	50.00	50.00
	CI-PICA	/train + /dev	53.03	50.75	48.99
	QDA		85.00	56.90	65.30
	LDA		53.03	50.75	48.99
	KNN		94.29	89.64	62.89
	RNF		100.00	99.31	64.34
	KNM		53.30	51.34	48.86

Figure 1: Data Table

B. PLOTTING THE ORIGINAL DATA

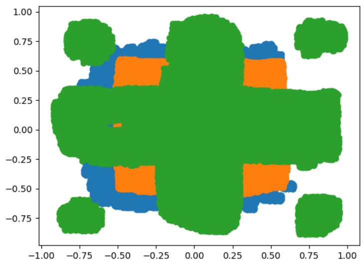

Figure 2: Training Set 8

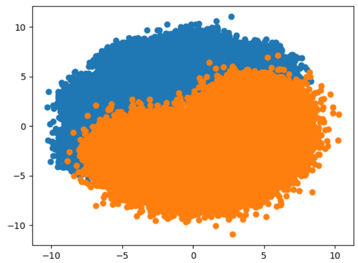

Figure 3: Training Set 9

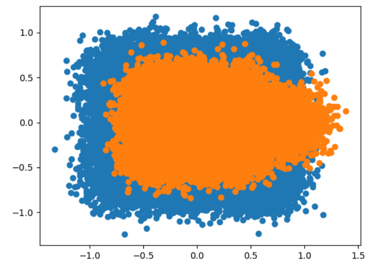

Figure 4: Training Set 10

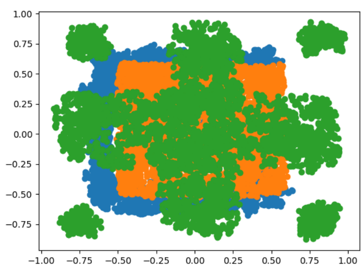

Figure 5: Development Set 8

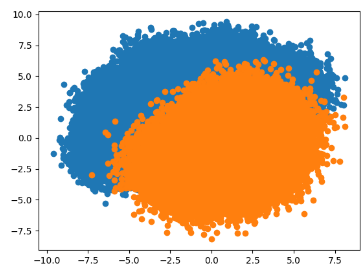

Figure 6: Development Set 9

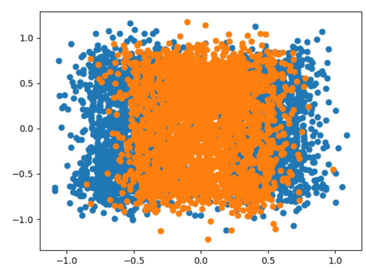

Figure 7: Development Set 10

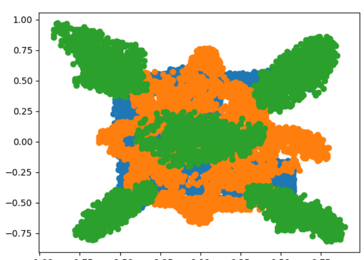

Figure 8: Evaluation Set 8

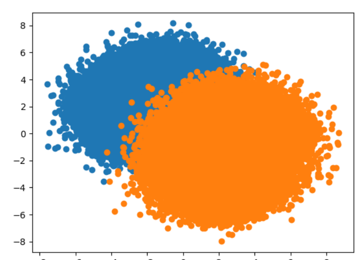

Figure 9: Evaluation Set 9

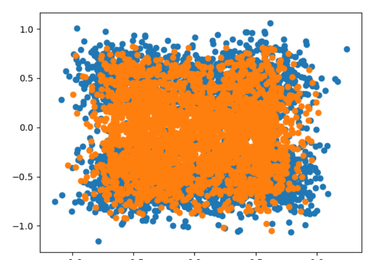

Figure 10: Evaluation Set 10

C. DECISION SURFACES

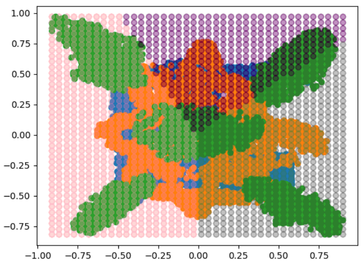

Figure 11: Set 8 PCA

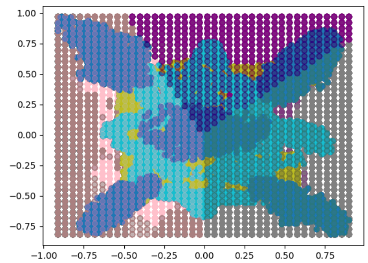

Figure 12: Set 9 PCA

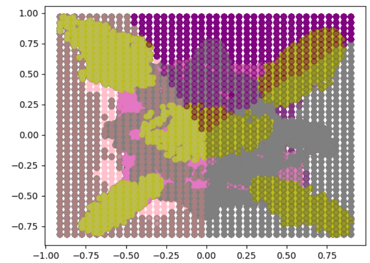

Figure 13: Set 10 PCA

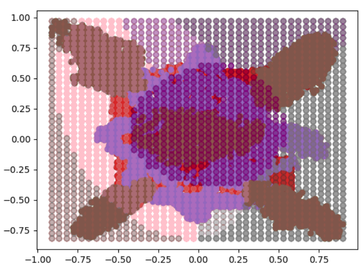

Figure 14: Set 8 QDA

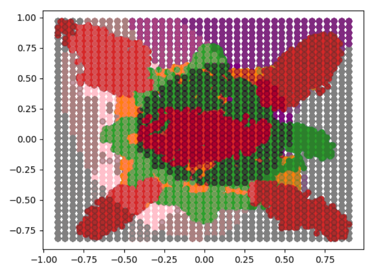

Figure 15: Set 9 QDA

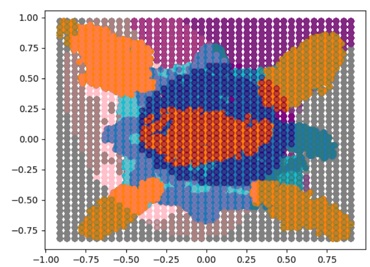

Figure 16: Set 10 QDA

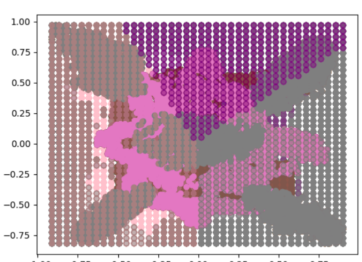

Figure 17: Set 8 LDA

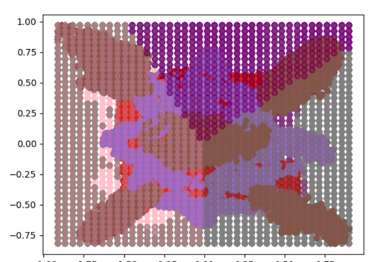

Figure 18: Set 9 LDA

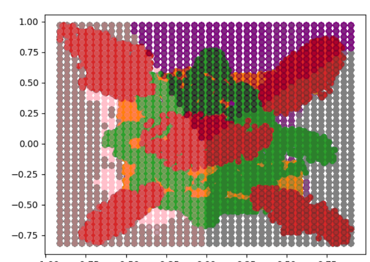

Figure 19: Set 10 LDA

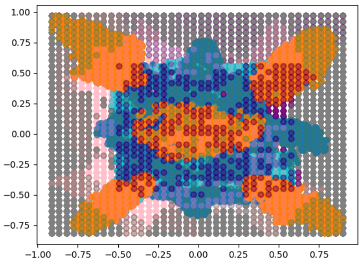

Figure 20: Set 8 KNN

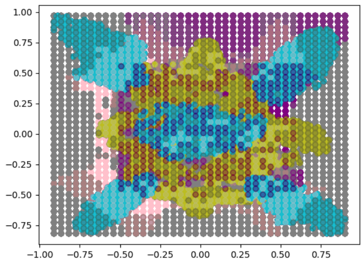

Figure 21: Set 9 KNN

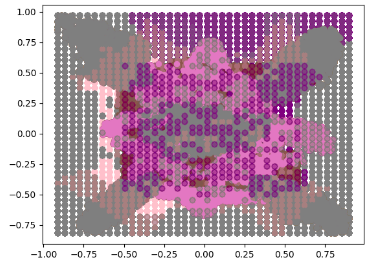

Figure 22: Set 10 KNN

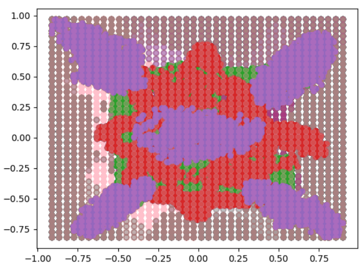

Figure 23: Set 8 KNM

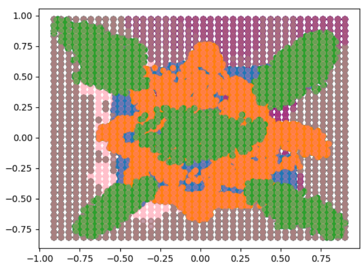

Figure 24: Set 9 KNM

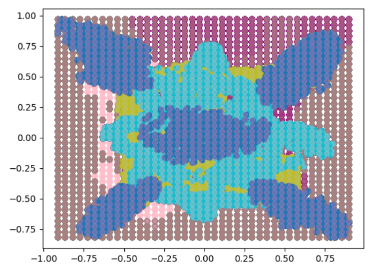

Figure 25: Set 10 KNM

D. KNN AND KNM PLOTS

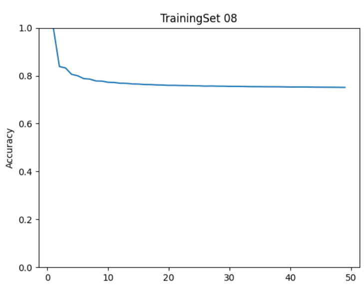

Figure 26: Training Set 8

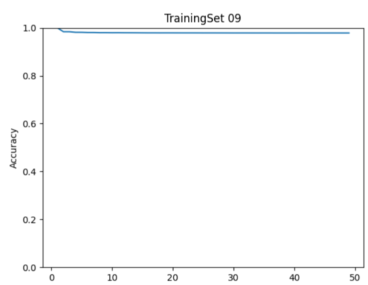

Figure 27: Training Set 9

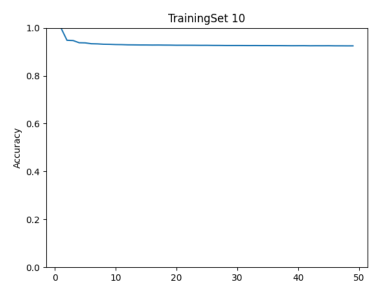

Figure 28: Training Set 10

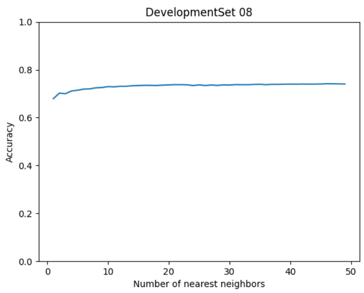

Figure 29: Development Set 8

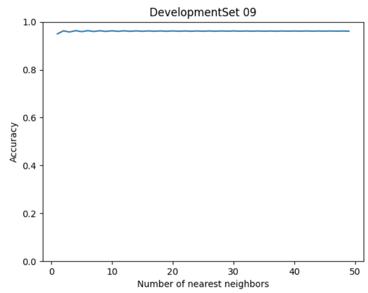

Figure 30: Development Set 9

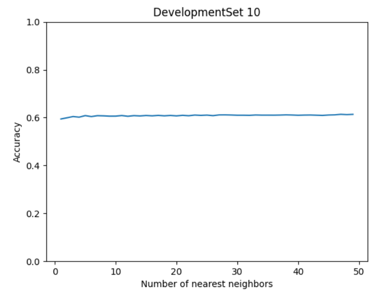

Figure 31: Development Set 10

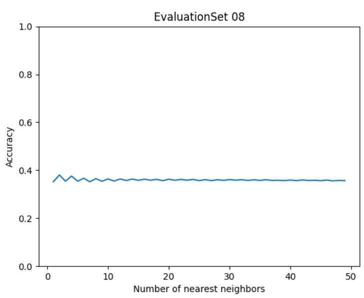

Figure 32: Evaluation Set 8

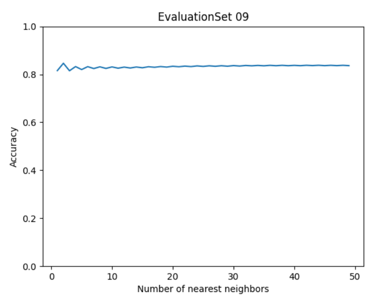

Figure 33: Evaluation Set 9

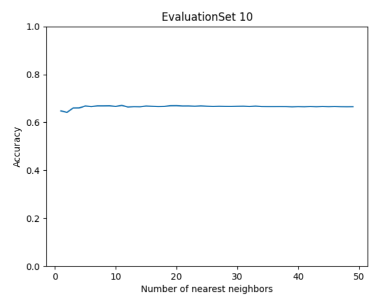

Figure 34: Evaluation Set 10

E. SUMMARY

For dataset 8, we can see that our best result for the eval data was QDA. By calculating the mean and covariance for each dataset, we can see that even though it didn't do as well on the training and dev set, it generalized the data so although it wasn't able to separate the classes well and squash everything into matrixes well, it was consistent which is important.

For dataset 9, we can see that QDA, PCA, and LDA did really well which could hint that the classes are separated clearly on a dimensionally reduced plane. KNN, RNF, and KNM didn't do as well which could possibly be indicative of the lack of normalization being a problem.

For dataset 10, we can see that KNN, RNF, and QDA perform the best. This is indicative of their being a large amount of variance. Since LDA didn't do as well as QDA we can see that the variance of data might be higher which shows that QDA might be better.

In summary, we can see that overall QDA performs the best.