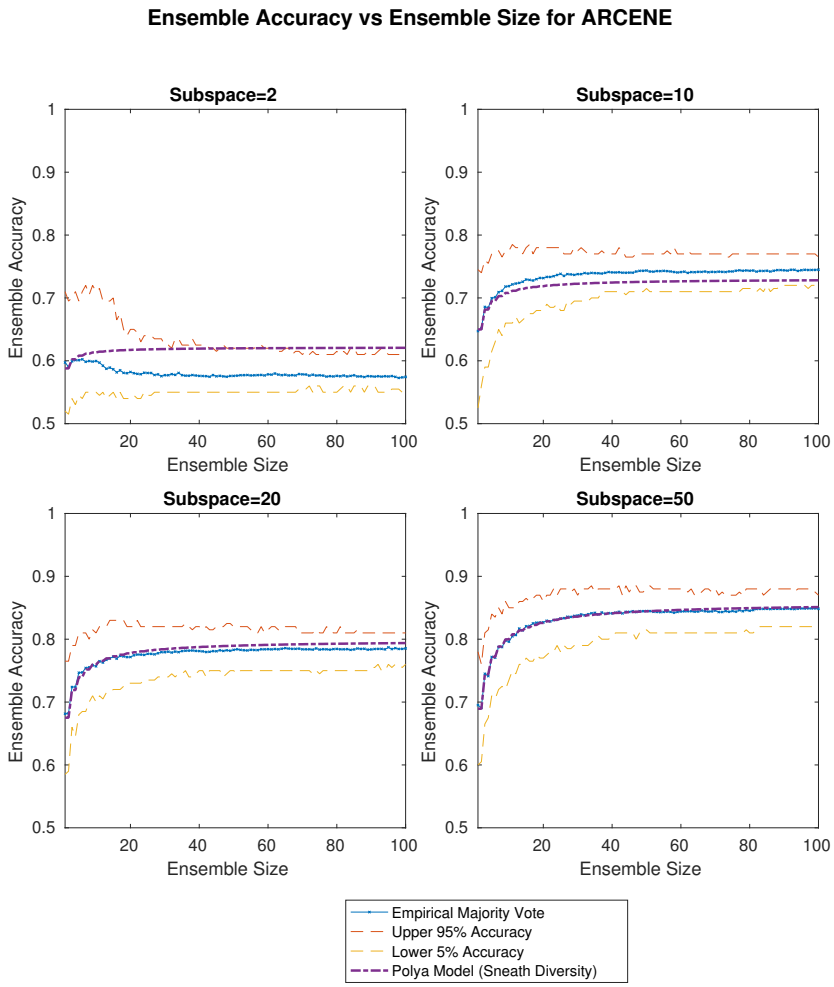

Figure 1: Average Majority vote ensemble accuracy for ARCENE using LDA-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe the non-monotonic behavior in the empirical accuracy when the number of projection dimensions is small. Also observe that at larger number of projection dimensions, our model predicts the average ensemble accuracy almost exactly.

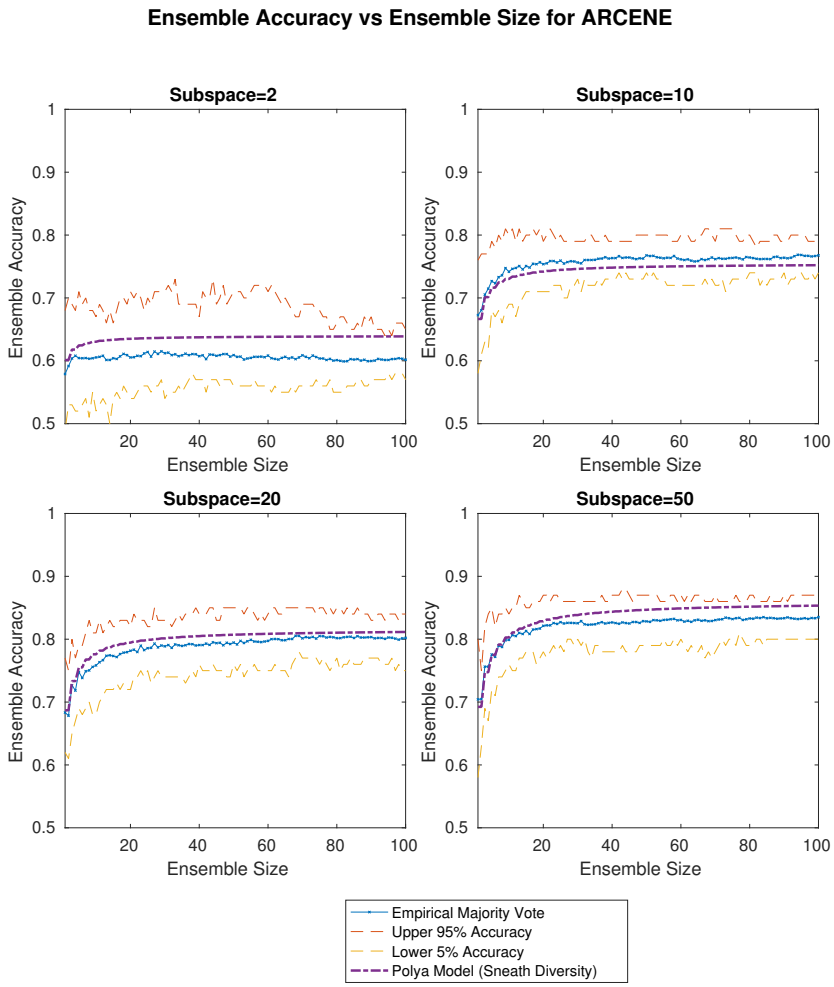

Figure 2: Average Majority vote ensemble accuracy for ARCENE using LDA-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe the non-monotonic behavior in the empirical accuracy when the number of projection dimensions is small. Also observe that at larger number of projection dimensions, our model predicts the average ensemble accuracy almost exactly.

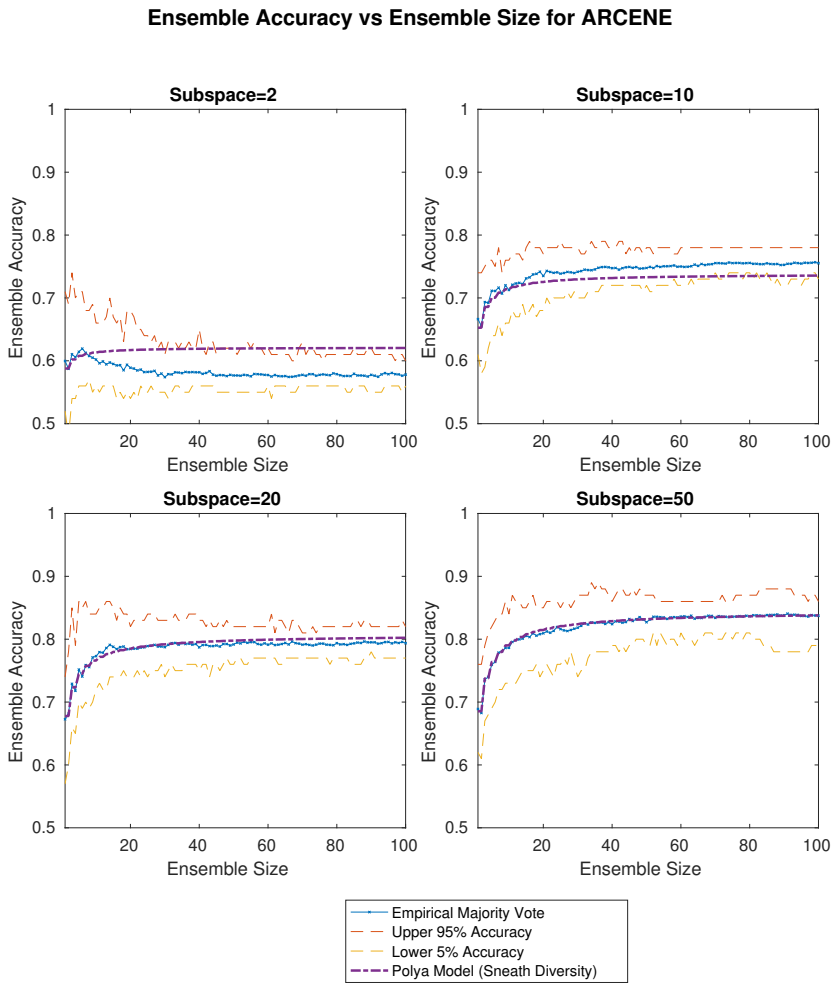

Figure 3: Average Majority vote ensemble accuracy for ARCENE using SVM-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe the non-monotonic behavior in the empirical accuracy when the number of projection dimensions is small. Also observe that at larger number of projection dimensions, our model predicts the average ensemble accuracy almost exactly..

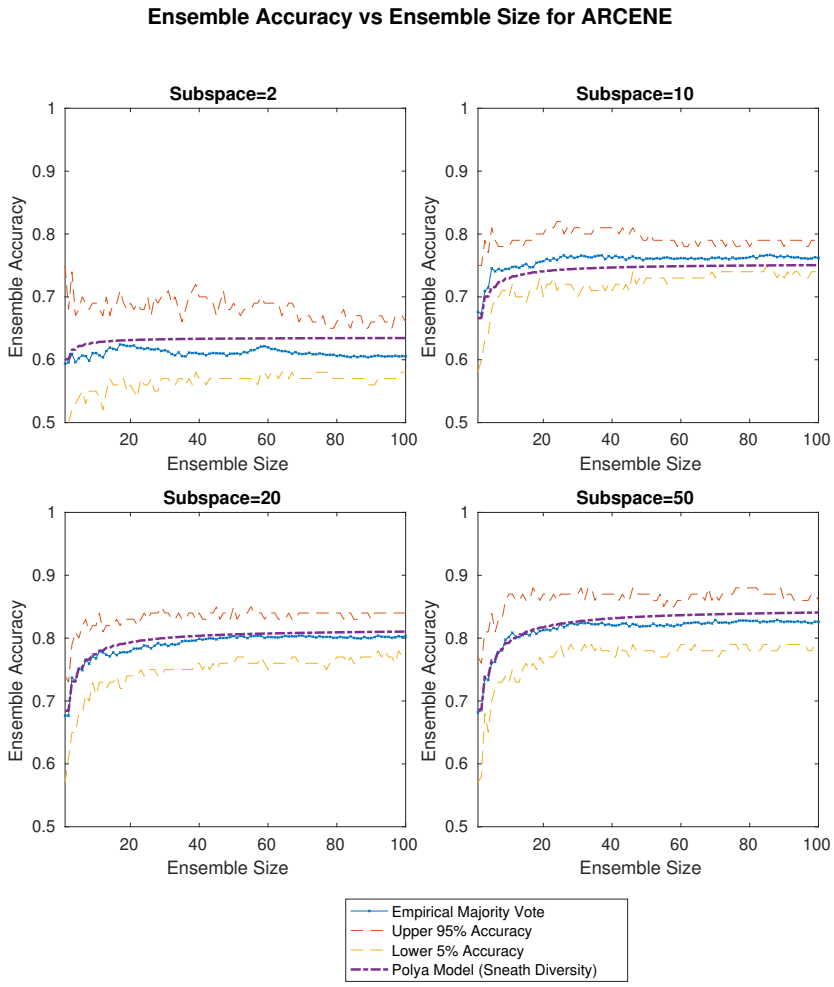

Figure 4: Average Majority vote ensemble accuracy for ARCENE using SVM-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe the non-monotonic behavior in the empirical accuracy when the number of projection dimensions is small. Observe that our model predicts the average ensemble accuracy almost exactly.

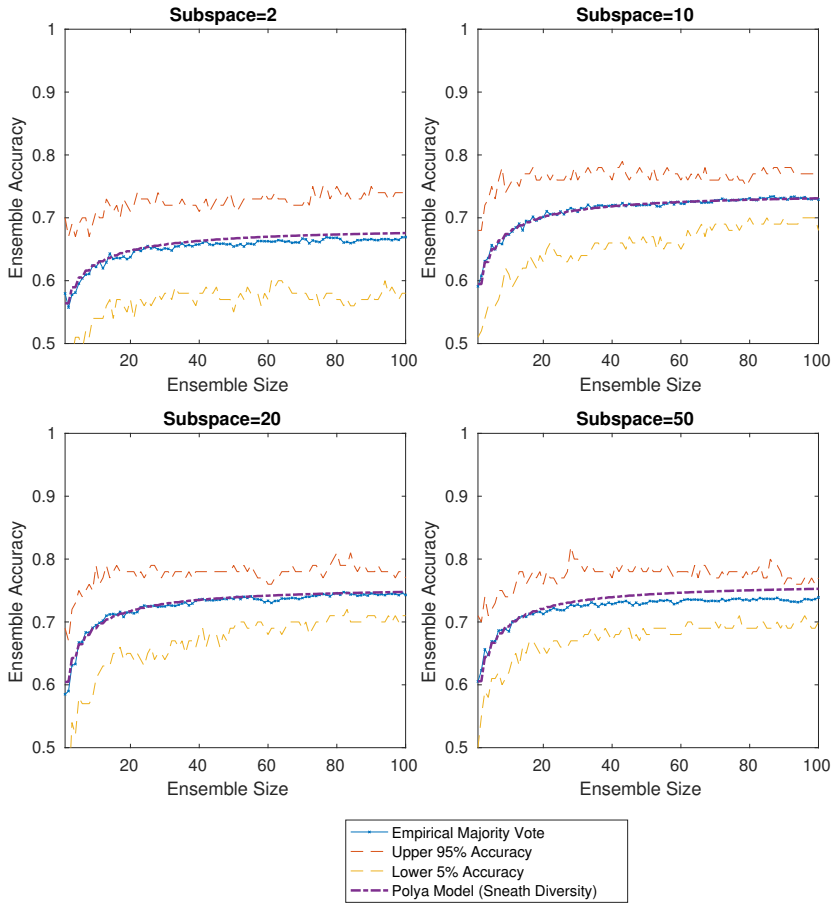

Figure 5: Average Majority vote ensemble accuracy for ARCENE using random forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

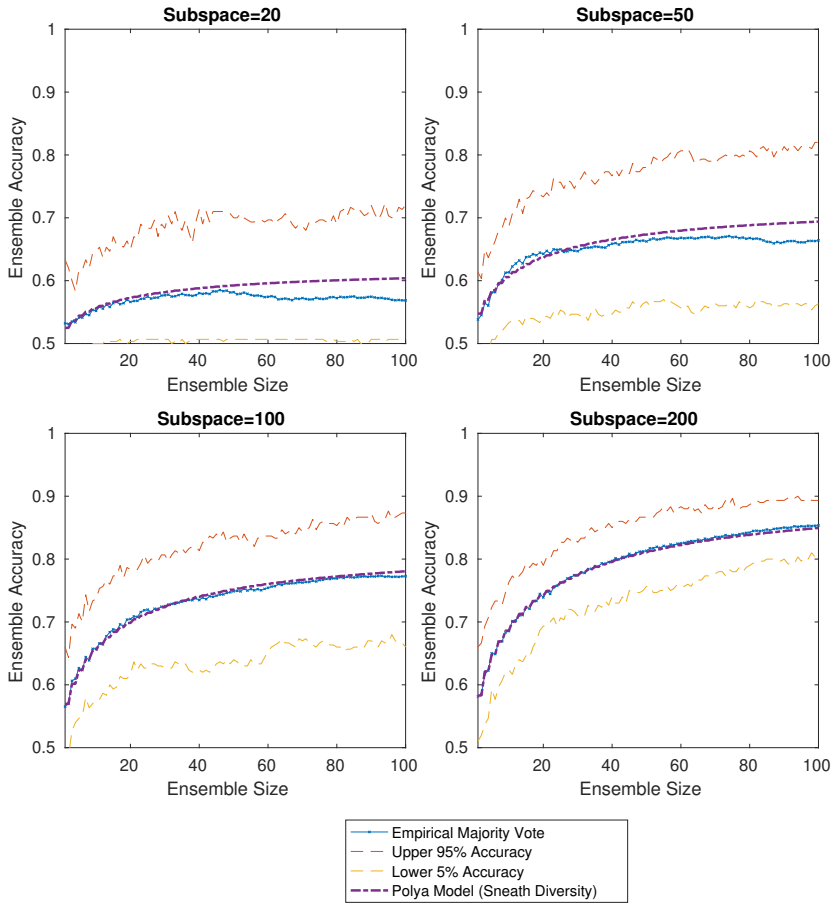

Figure 6: Average Majority vote ensemble accuracy for DEXTER using LDA-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy improves as the number of projection dimensions increases.

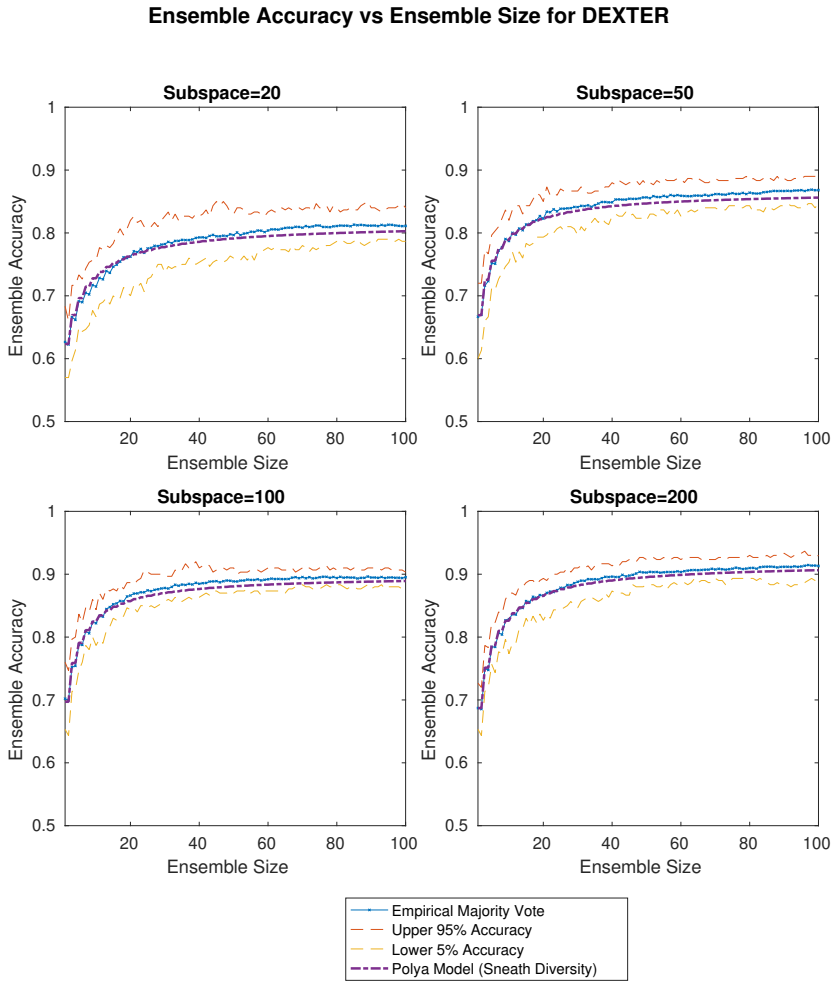

Figure 7: Average Majority vote ensemble accuracy for DEXTER using LDA-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

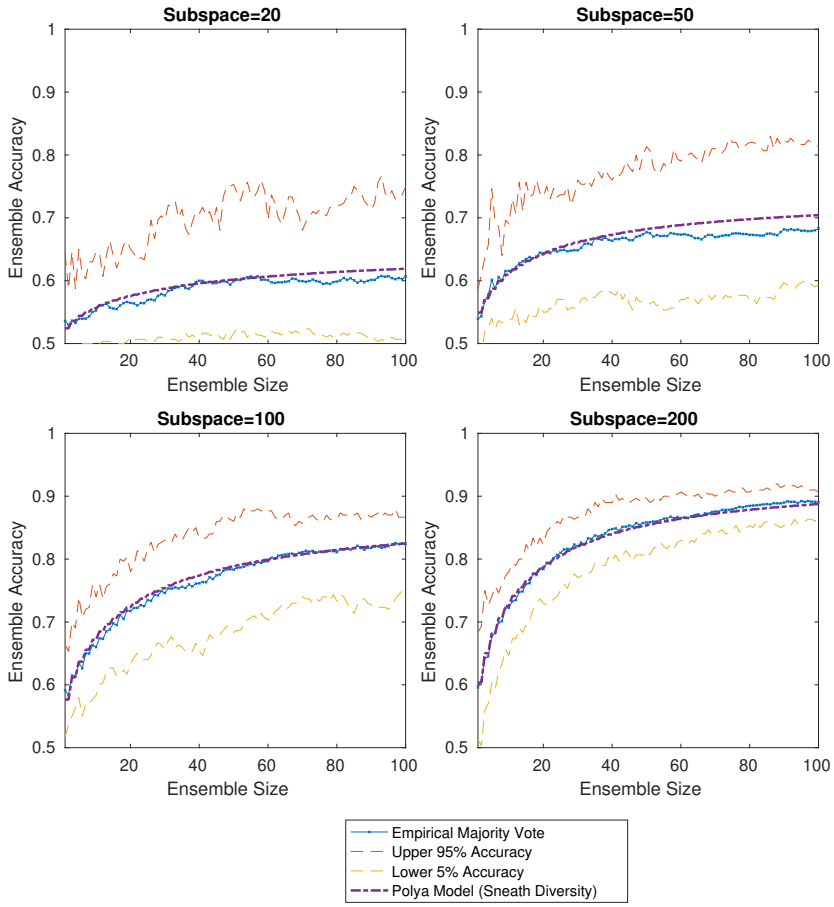

Figure 8: Average Majority vote ensemble accuracy for DEXTER using SVM-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

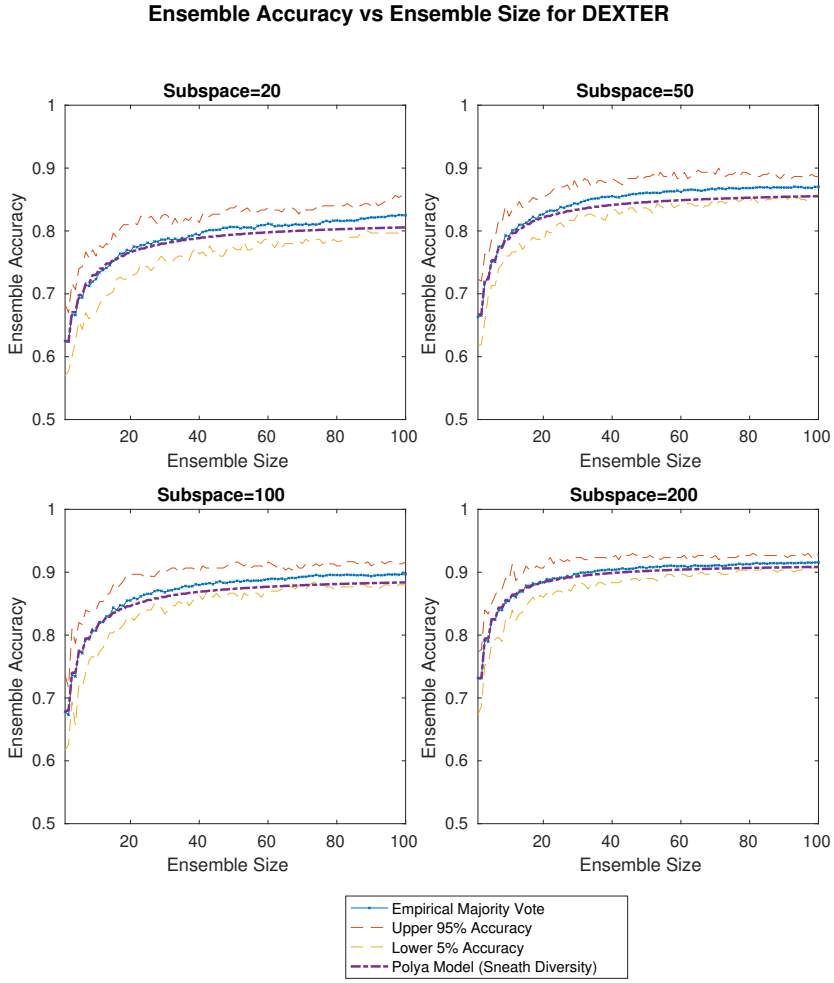

Figure 9: Average Majority vote ensemble accuracy for DEXTER using SVM-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

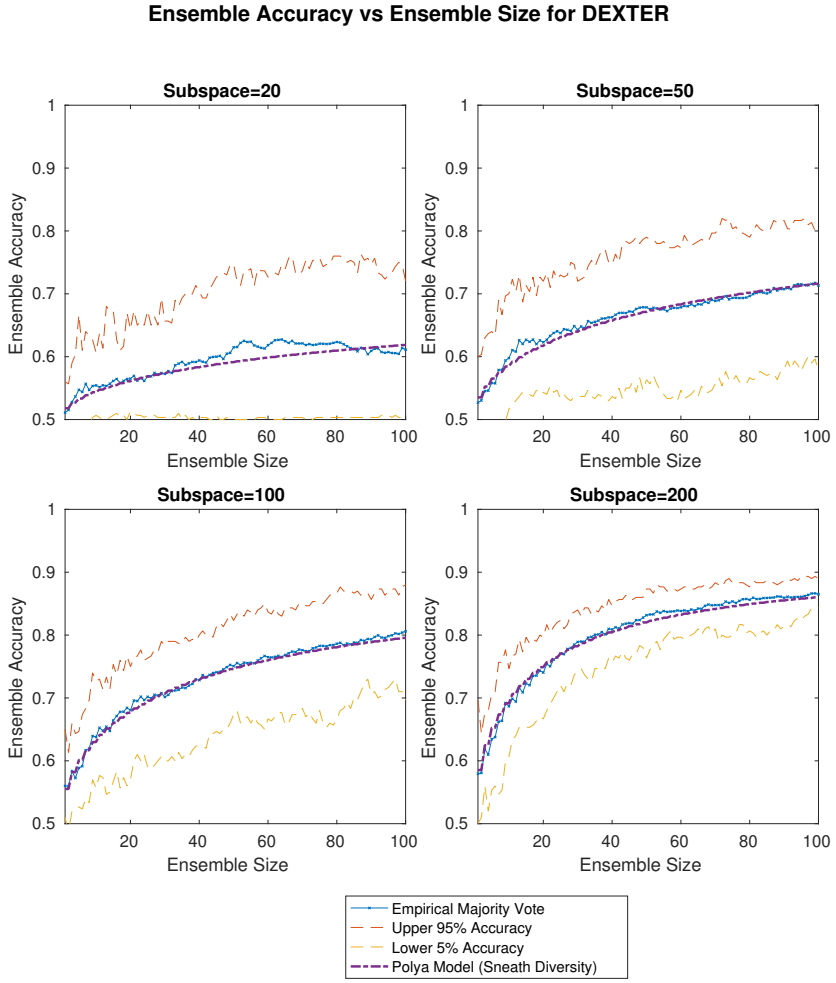

Figure 10: Average Majority vote ensemble accuracy for DEXTER using Random Forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

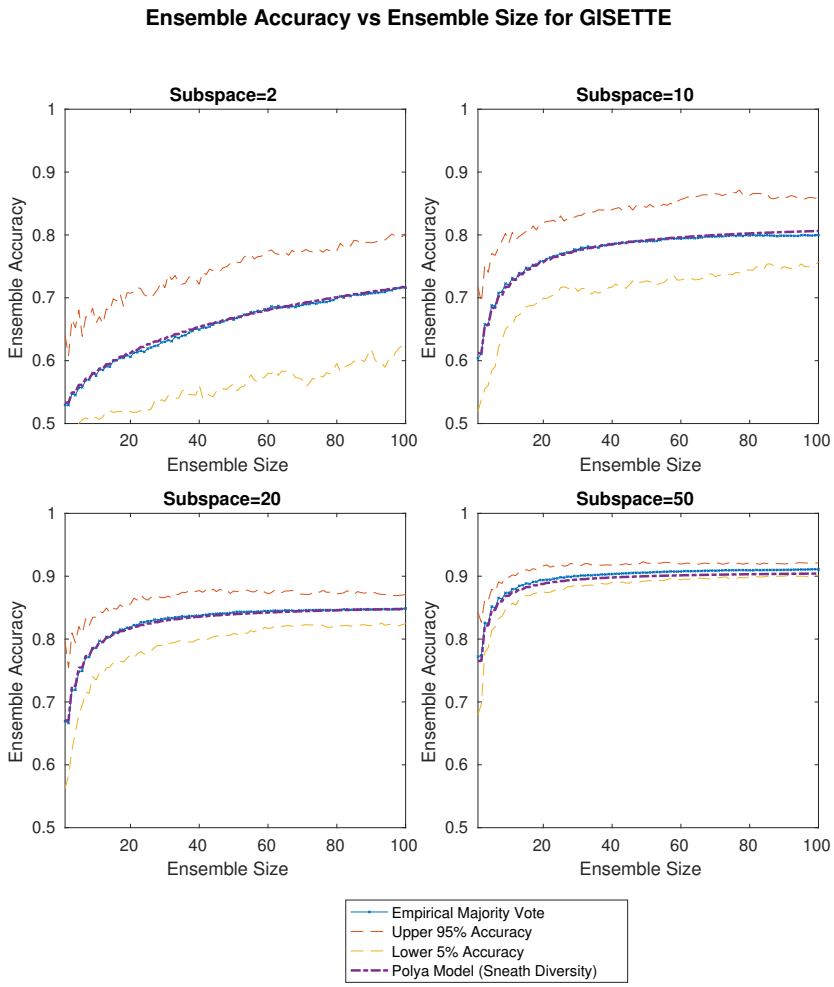

Figure 11: Average Majority vote ensemble accuracy for GISETTE using LDA-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

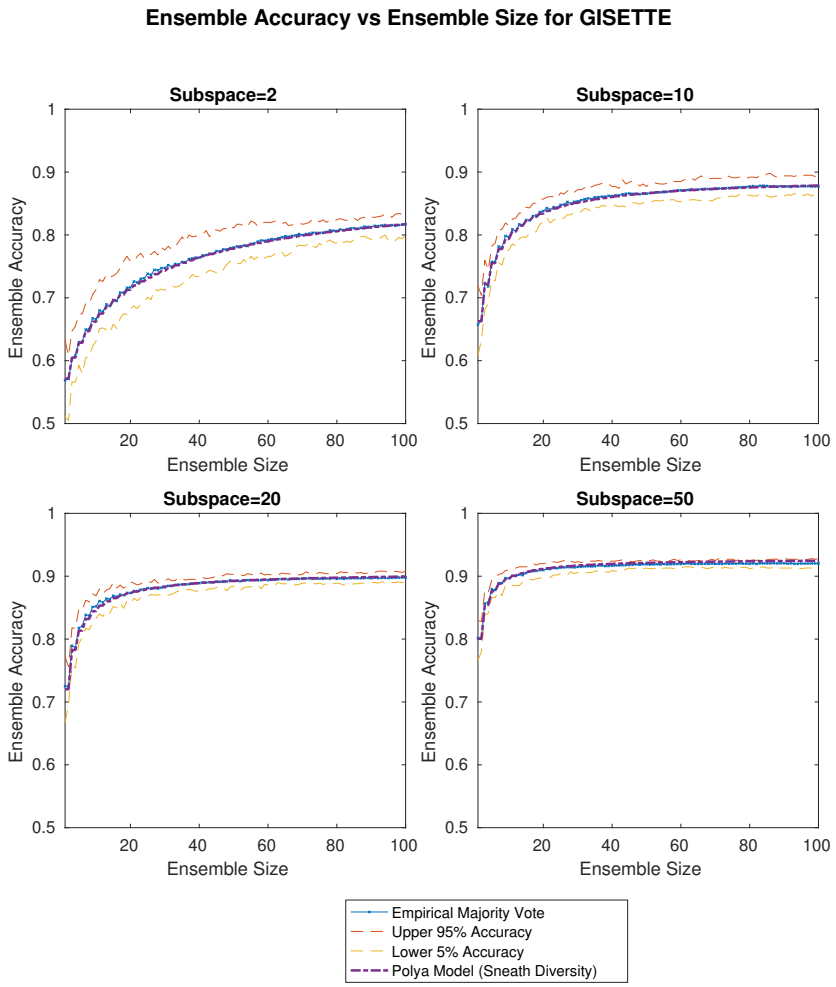

Figure 12: Average Majority vote ensemble accuracy for GISETTE using LDA-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

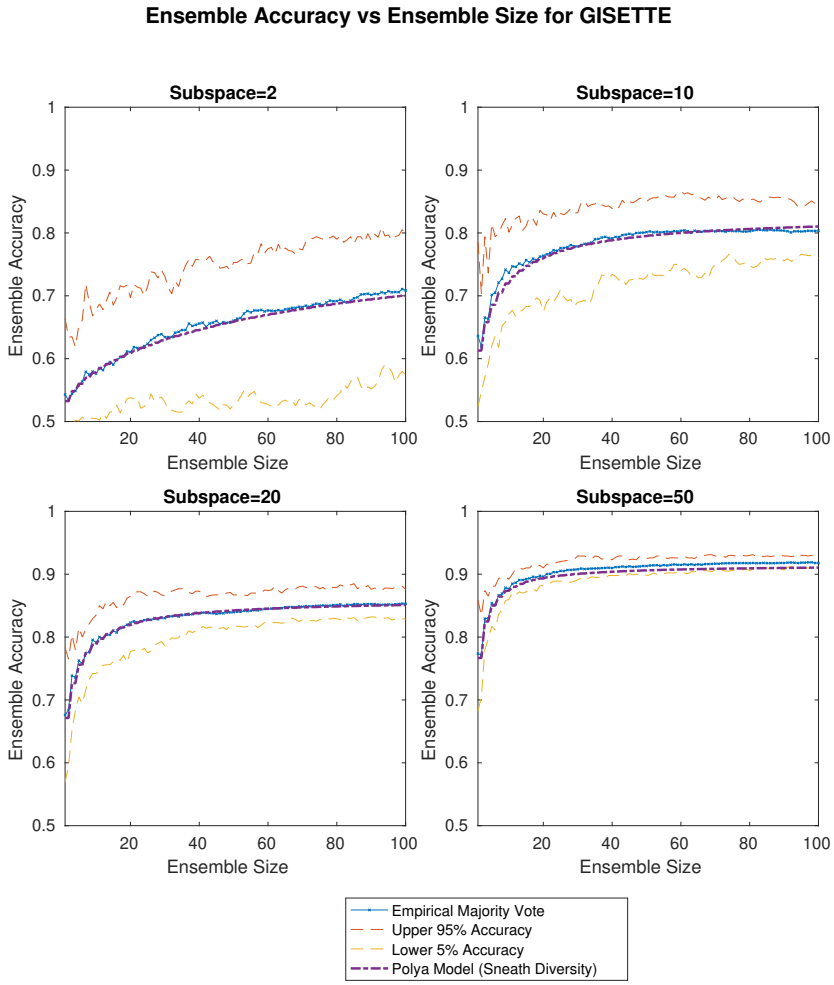

Figure 13: Average Majority vote ensemble accuracy for GISETTE using SVM-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

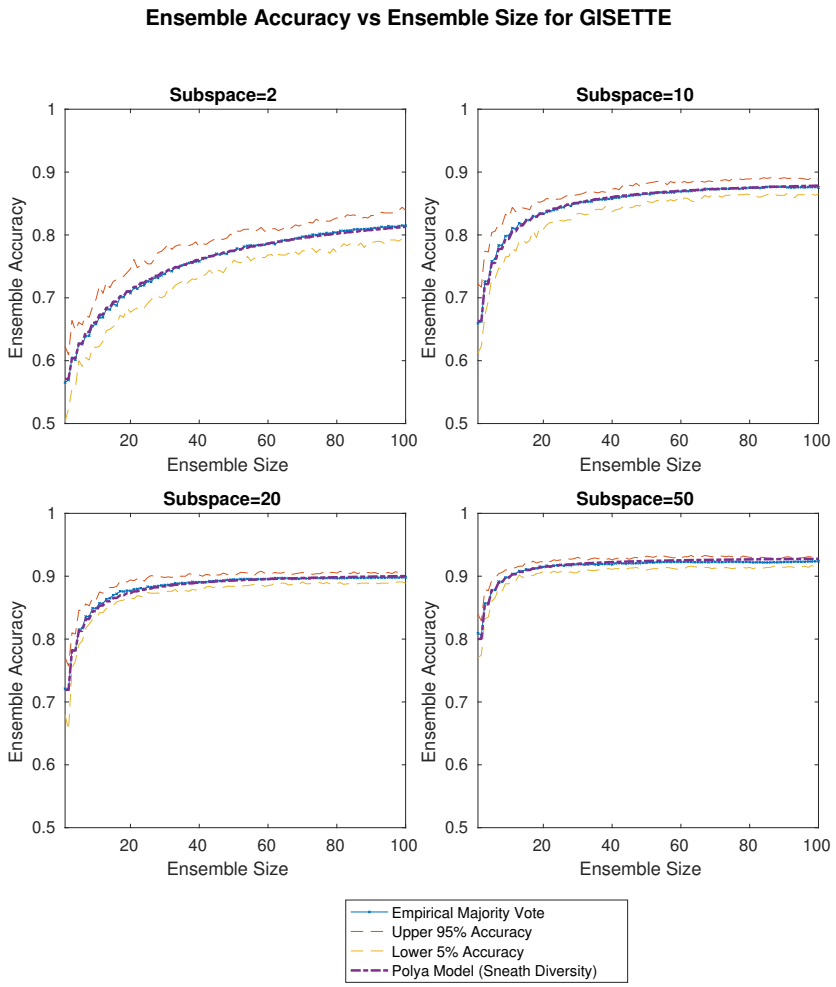

Figure 14: Average Majority vote ensemble accuracy for GISETTE using SVM-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

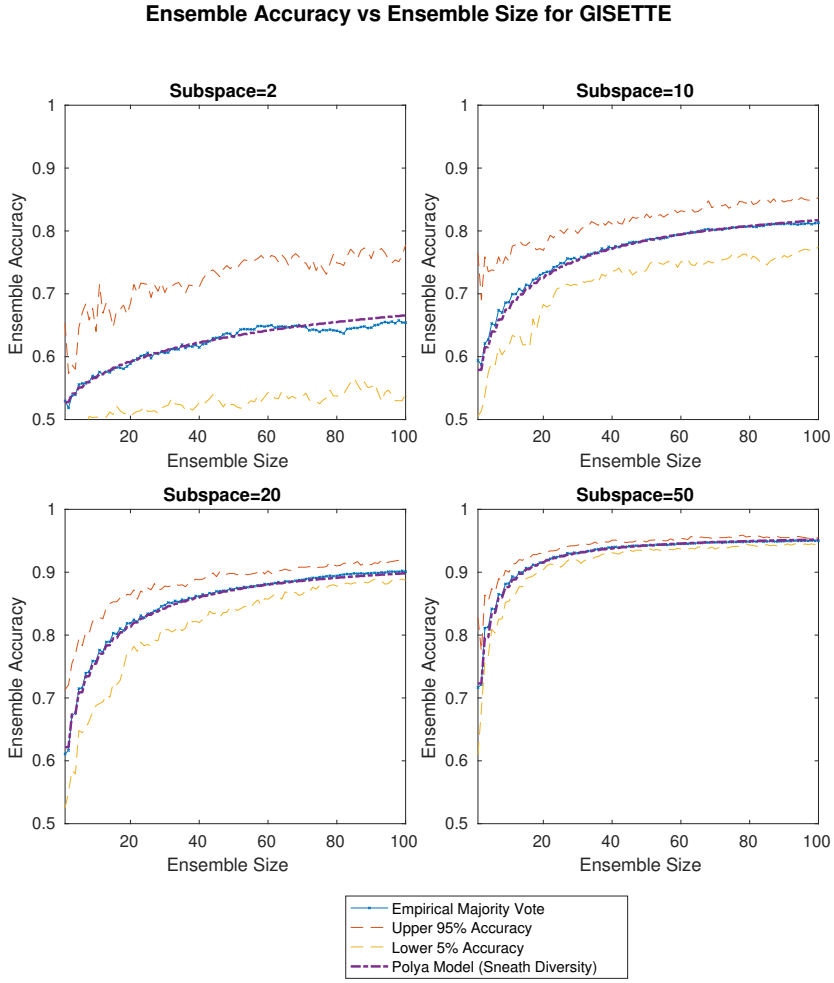

Figure 15: Average Majority vote ensemble accuracy for GISETTE using Random Forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

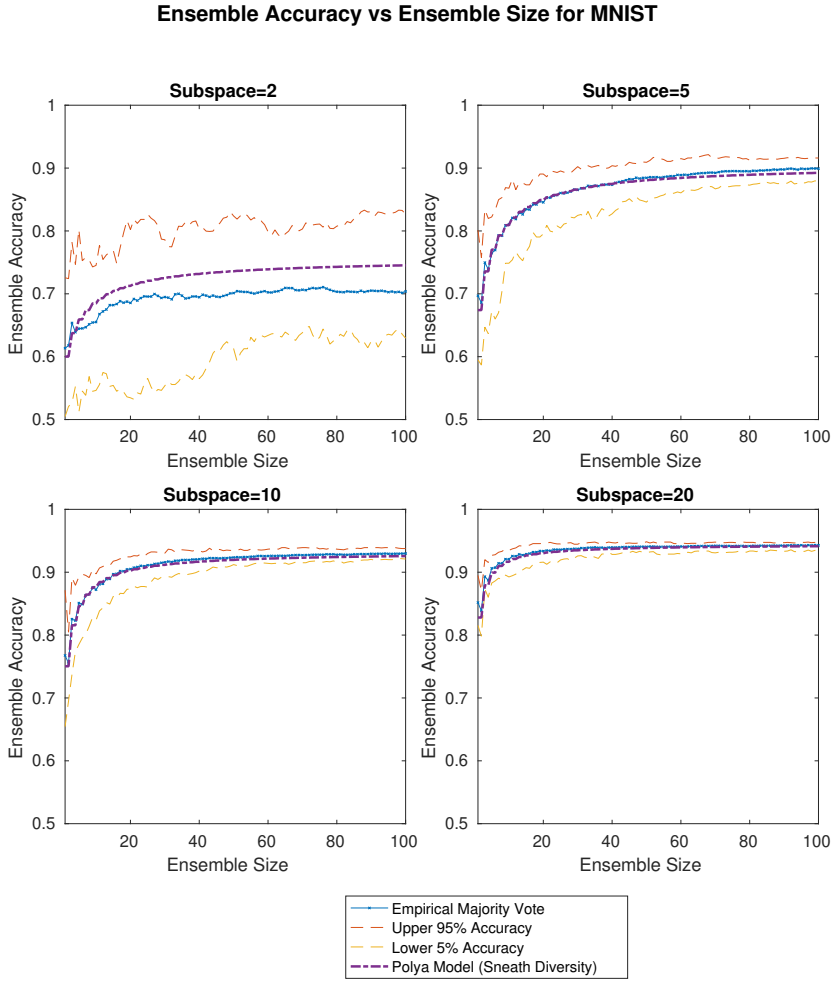

Figure 16: Average Majority vote ensemble accuracy for MNIST using LDA-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy improves as the number of projection dimensions increases.

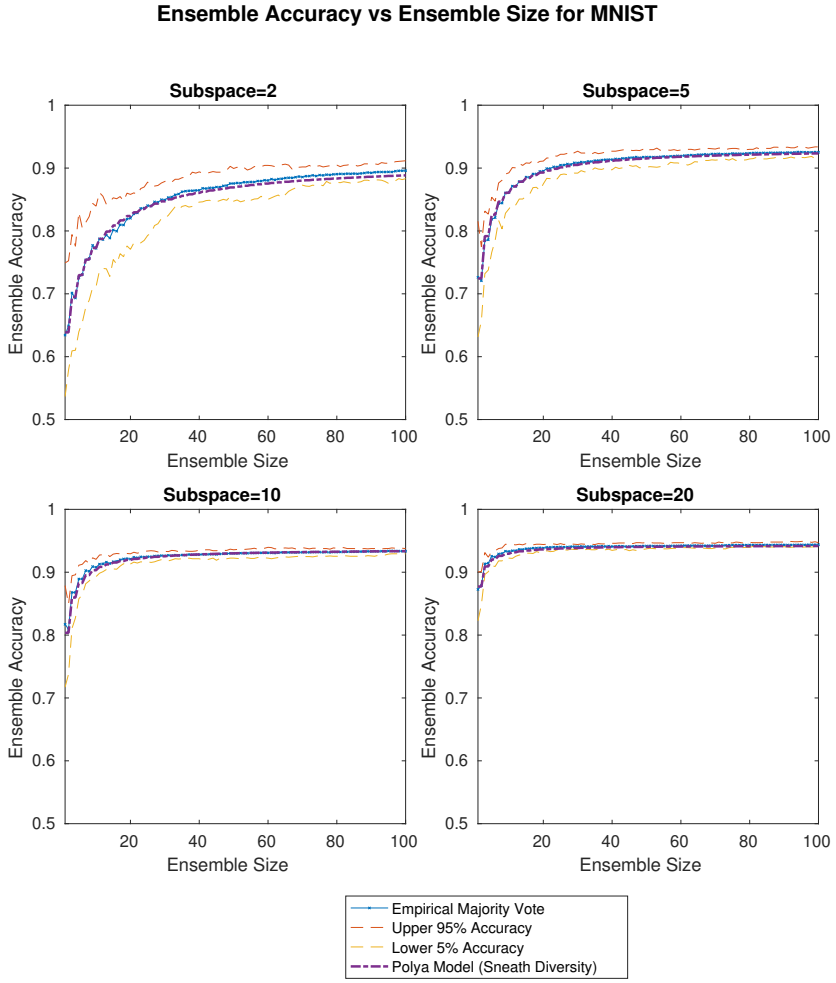

Figure 17: Average Majority vote ensemble accuracy for MNIST using LDA-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

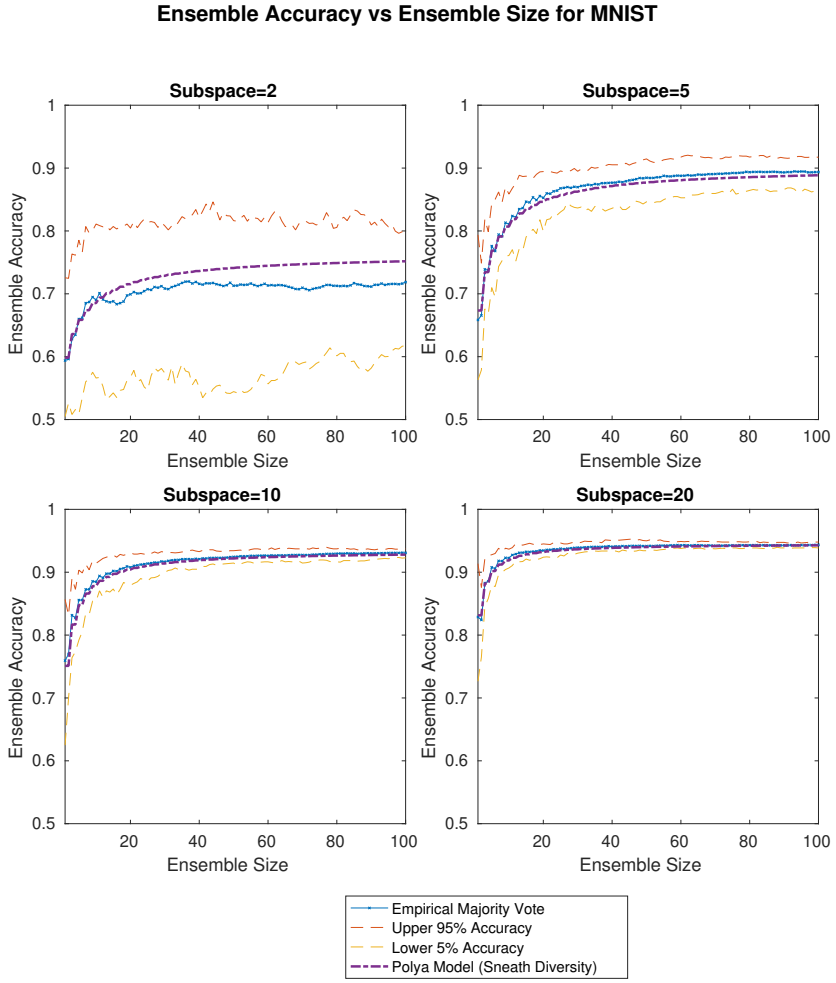

Figure 18: Average Majority vote ensemble accuracy for MNIST using SVM-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy better as the number of projection dimensions increases.

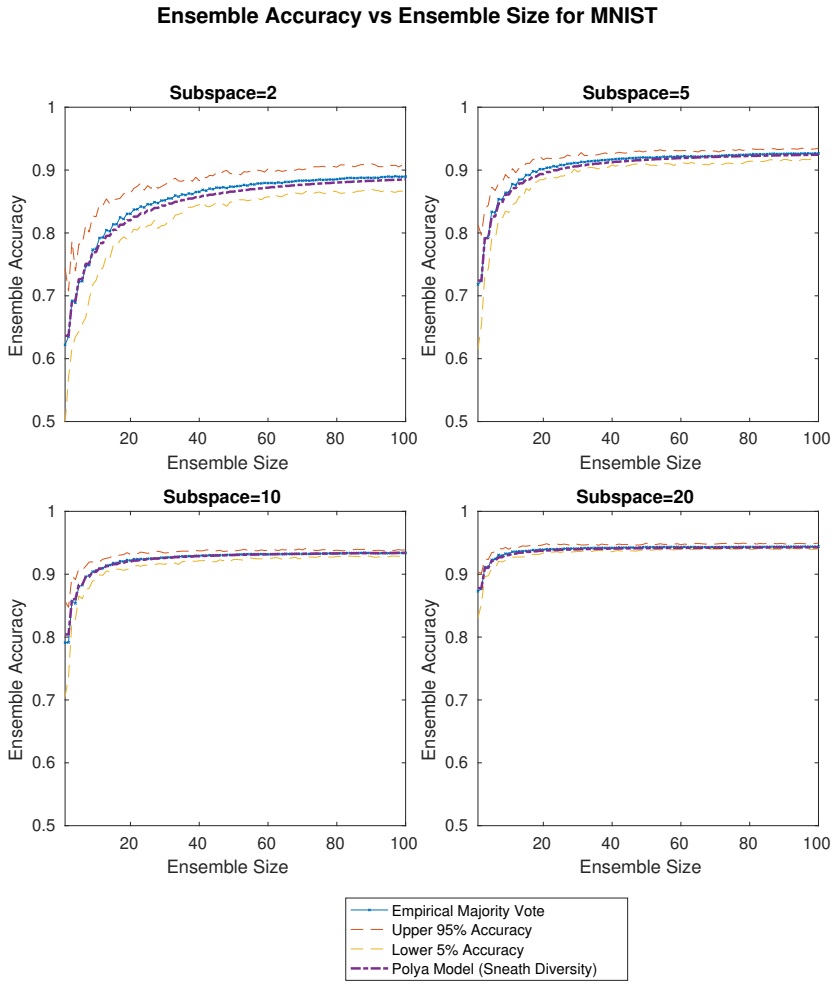

Figure 19: Average Majority vote ensemble accuracy for MNIST using SVM-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

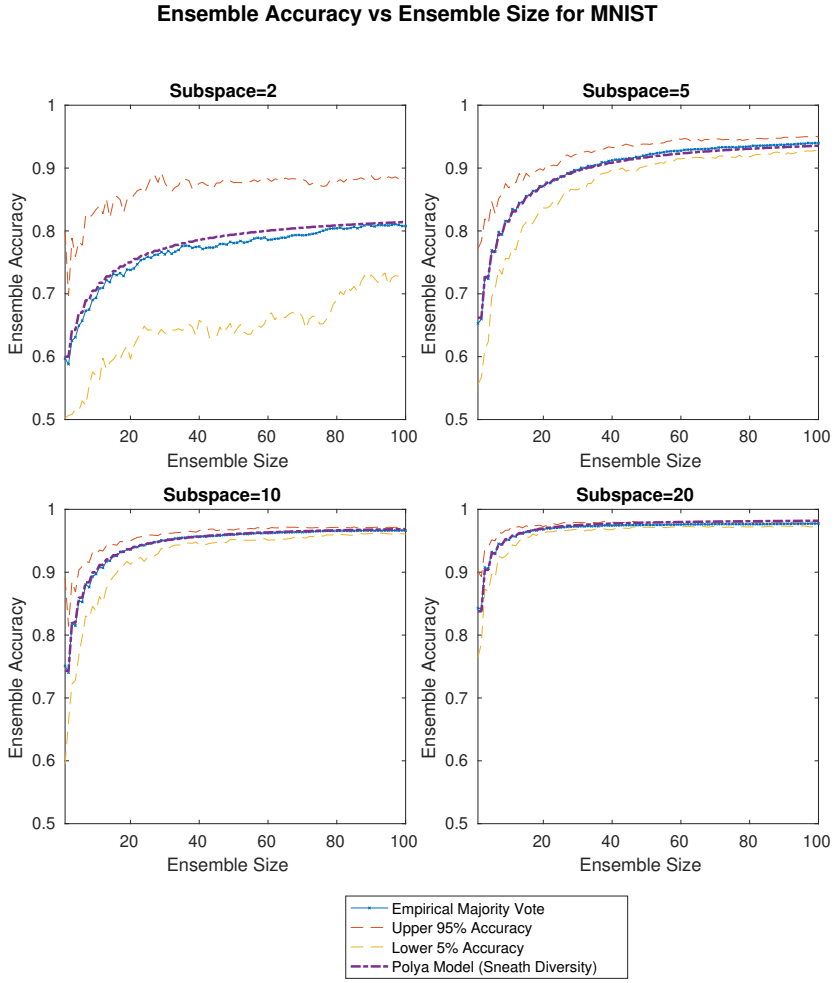

Figure 20: Average Majority vote ensemble accuracy for MNIST using Random Forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

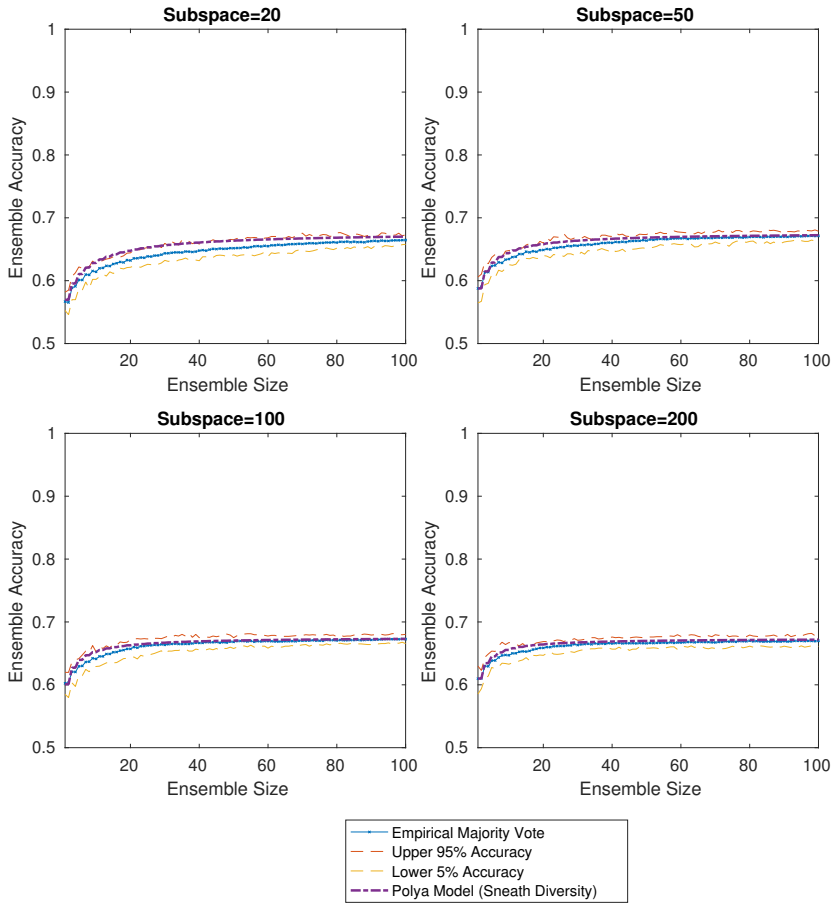

Figure 21: Average Majority vote ensemble accuracy for GTZAN using LDA-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

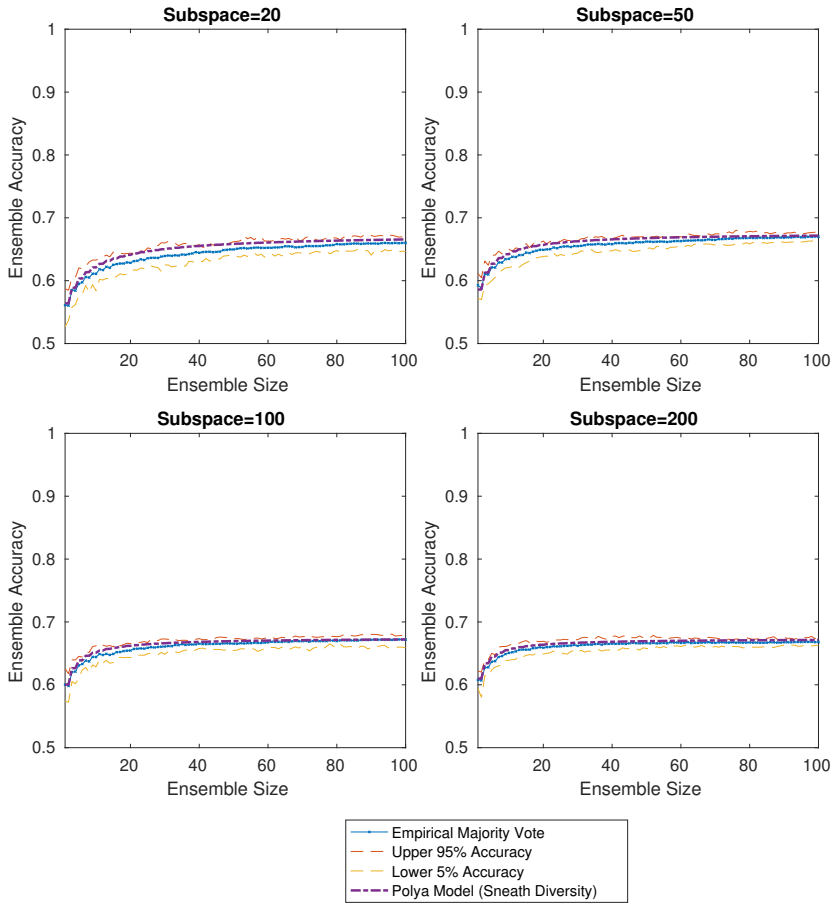

Figure 22: Average Majority vote ensemble accuracy for GTZAN using LDA-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

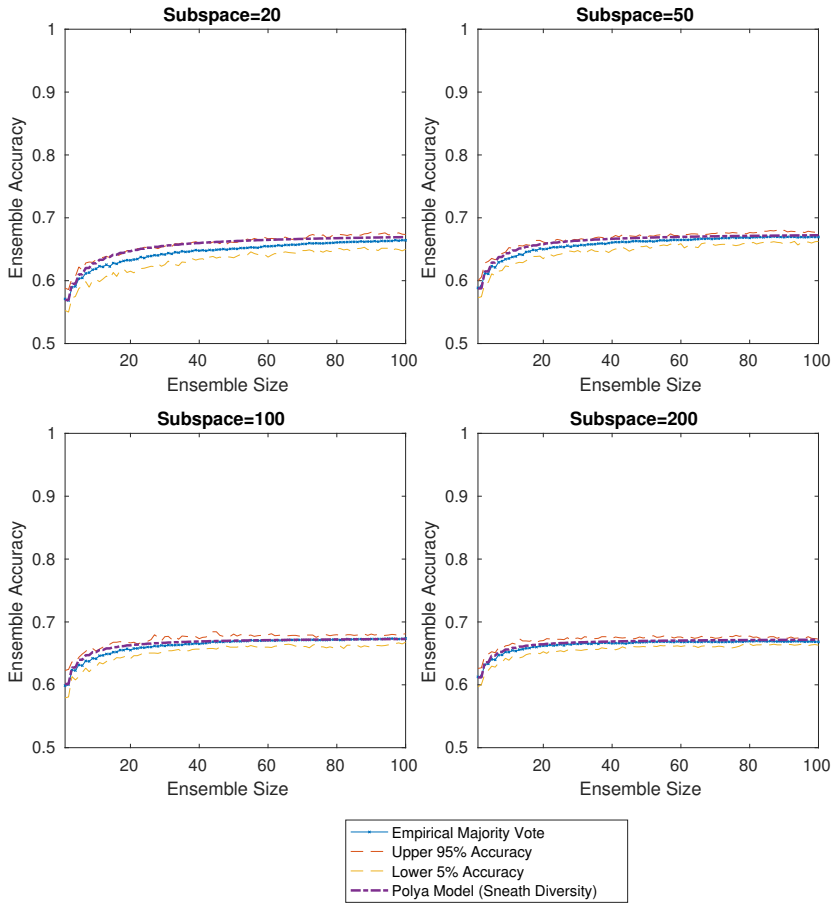

Figure 23: Average Majority vote ensemble accuracy for GTZAN using SVM base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

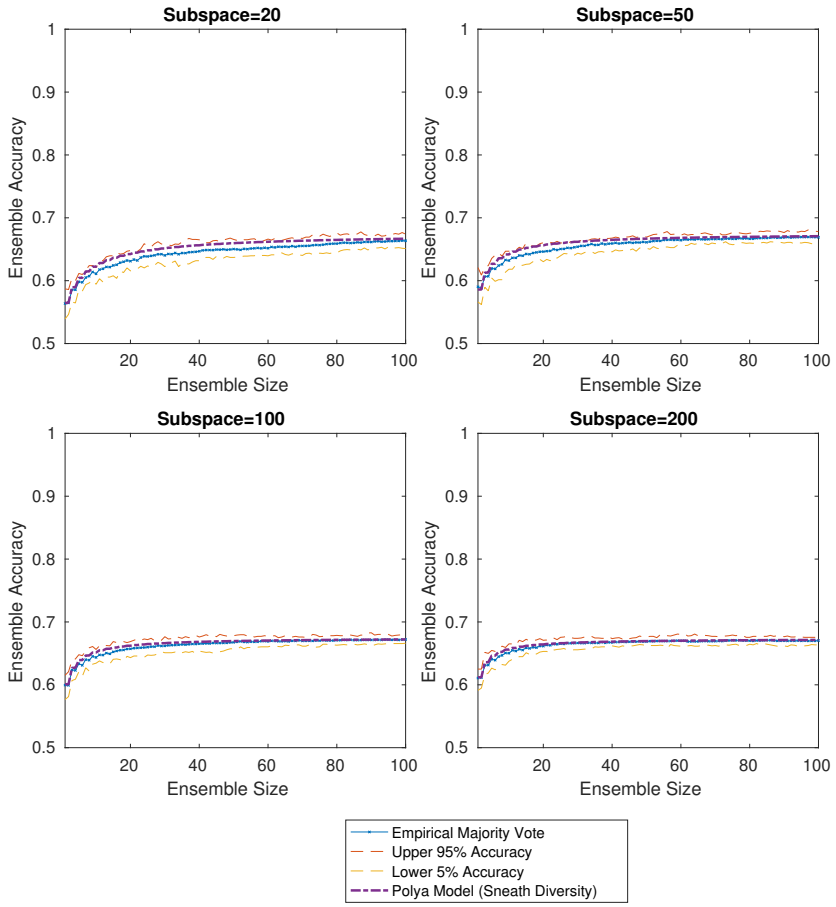

Figure 24: Average Majority vote ensemble accuracy for GTZAN using SVM-RP base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

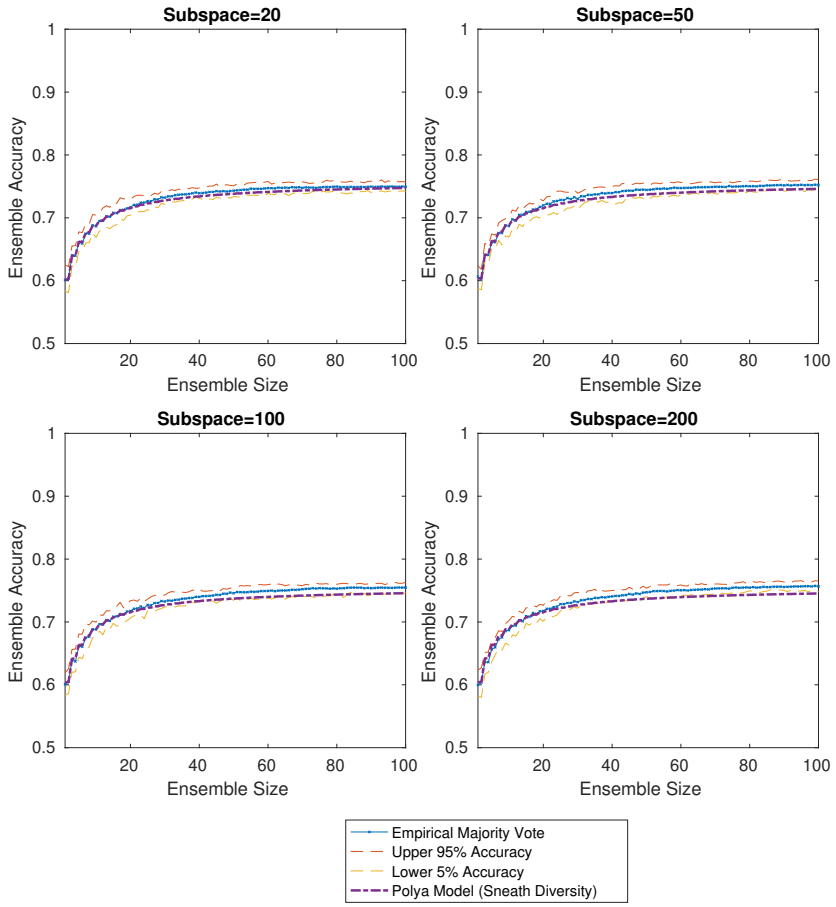

Figure 25: Average Majority vote ensemble accuracy for GTZAN using Random Forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

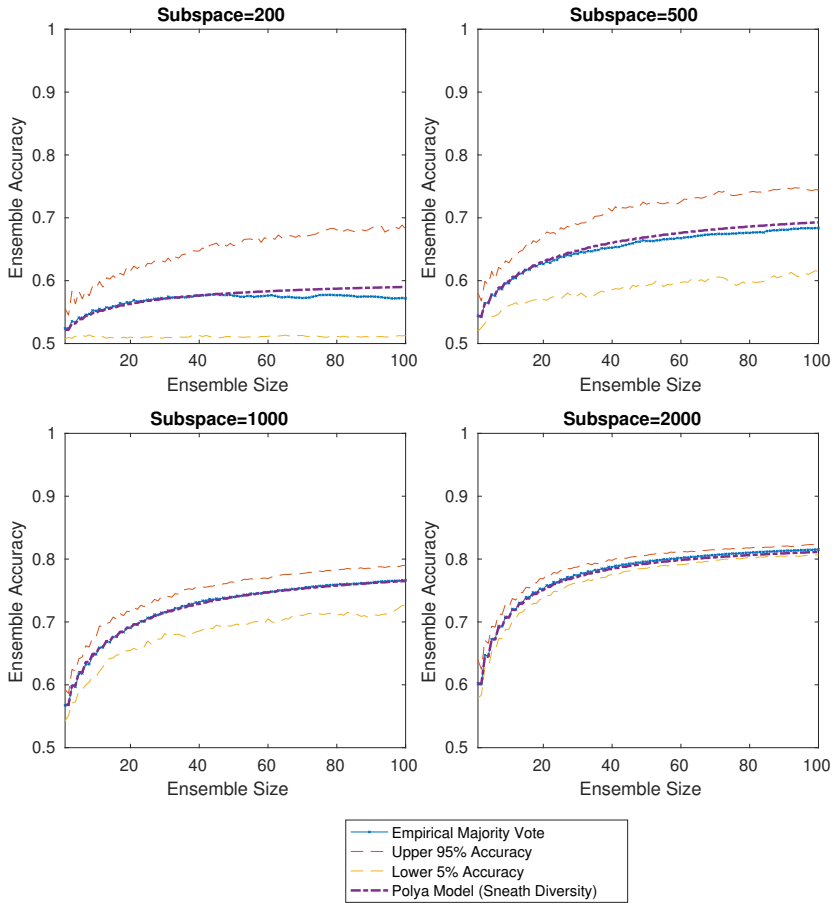

Figure 26: Average Majority vote ensemble accuracy for IMDB using SVM-RS base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.

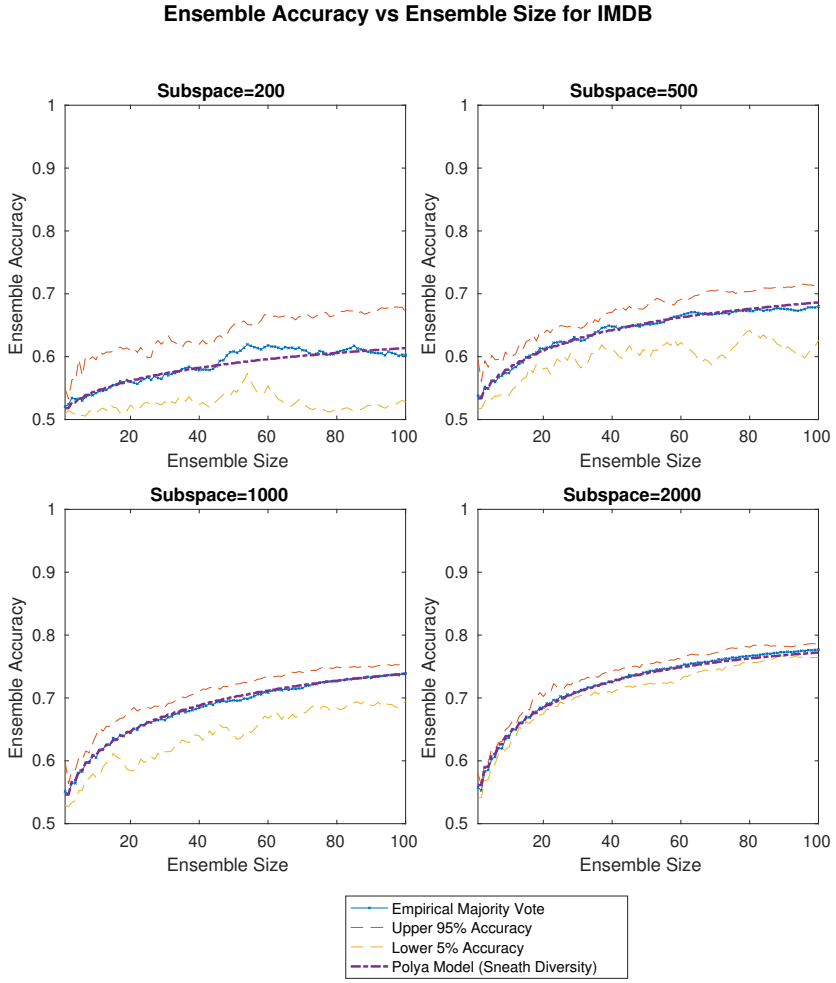

Figure 27: Average Majority vote ensemble accuracy for IMDB using Random Forest base classifiers as modelled by a Polya-Eggenberger distribution vs ensemble member size for various datasets on different base classifiers. The upper and lower ticks are bootstrap confidence intervals for the upper-95% and lower-5% empirical accuracy. Observe that our model predicts the average ensemble accuracy almost exactly.