CS5644 Fall 2018

Seunghyuk Baek

Sbaek44@vt.edu

Homework #2

Q1. Mean and standard deviation of decision tree classifier and Bayes classifier. The data was Congressional vote record contains three values (y for yes, n for no and ? for neither y or n). First, the data was converted into 1 for 'y' and 0 for 'n'. The data was sorted to 3 different scenarios. 1. All instances with '?' values removed. 2. All '?' were replaced with 2. 3. All missing values were replaced with the most frequent value.

	Decision Tree			Bayes		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Remove Instance Mean	0.946	0.944	0.944	0.92	0.914	0.914
Remove Instance STDev	0.034351	0.035071	0.035071	0.067823	0.073007	0.073007
Replace Impurity Mean	0.954	0.948	0.948	0.906	0.9	0.9
Replace Impurity STDev	0.015166	0.014832	0.014832	0.040373	0.048477	0.048477
Replace with Frequent Value Mean	0.954	0.95	0.95	0.906	0.9	0.9
Replace with Frequent Value STDev	0.015166	0.018708	0.018708	0.040373	0.048477	0.048477

Decision Tree			
	precision	recall	f1-score
mean	0.951333	0.947333	0.947333
std	0.021996	0.022824	0.022824
Bayes			
	precision	recall	f1-score
mean	0.910667	0.904667	0.904667
std	0.047879	0.053966	0.053966

2. Based on statistical metrics, it is better to choose decision tree classifier. Although the difference of mean between decision tree classifier and Bayes classifier, decision tree classifier shows less standard deviation.