Q1) Distinguish between Classification and Regression with the help of relevant scenarios.?

Classification vs Regression

Ans) Classification is a data mining function that assigns itms in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

> Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features.

Example of Regression:

Predictive percentage of precipitation .

Predictiion of diabetic patients.

Astronomy prediction.

Example of classification:

Who will voters vote for?

Who will win most matches?

Hidden Patterns.

Q2)Perform Naïve Bayes or Decision tree classification for new instance where (SSN = 123-46-4455, Test1= 85, Test2= 31 and Final= 30) Find Grade.

Ans) Decision tree: predict the class label – Bayesian classifier: statistical classifier; predict class membership probabilities.

Naïve Bayesian classifier:

- Simple classifier that assumes attribute independence
- Efficient when applied to large databases
- Comparable in performance to decision trees

 $P(H_i | X) = \underline{P(X | H_i) P(H_i)}$

P(X)

Q3

Ans)

Opening File

Adding filter (Numeric to Nominal) so Naïve Bayes could be applied.

🔇 Weka Explorer				- 0 ×
Preprocess Classify Cluster Associate Select attributes V	isualize			
Open file Open URL	Open DB Gene	unate Un	do Edit	Save
Filter				
Choose None				Apply Stop
Current relation		Selected attribute		
Relation: iris Instances: 150	Attributes: 5 Sum of weights: 150	Name: sepallength Missing: 0 (0%)	Distinct: 35	Type: Numeric Unique: 9 (6%)
Attributes		Statistic	Value	
		Minimum	4.3	
All None (Invert Dettern	Mean	5.843	
		StdDev	0.828	
No. Name				
1 sepallength				
3 petallength				
4 petalwidth				
5 🗌 class				
		Class: class (Nom)		Visualize Ali
			34	
		30	28	
			24	5
		16		
				7
Remove				
Status		4.3	6.1	7
ОК				Log 🛷 x0
📲 🗄 ڪ 🛤 💼 📤 🔇	S 🗷 🚾 🔇			へ 幅

Applying Naïve Bayes

Weka Explorer					- 0
Preprocess Classify Cluster Associ	iate Select attributes Visuali	lize			
ssifier					
Choose LibSVM -S 0 -K 2 -D 3 -G 0.0	-R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0).001 -P 0.1 -model "C:	\\Program Files\\Wel	a-3-8-4" -seed 1	
t options	Classifier output				
) Use training set	=== Run information) ===			
Supplied test set Set	Salaana		Nationa		
Cross-validation Folds 10	Relation: iris	Classifiers.bayes	.Maivebayes		
Percentage split % 66	Instances: 150 Attributes: 5				
More options	sepal	length			
	sepai petal	.width llength			
om) class	petal class	.width			
Start Stop	Test mode: split	: 66.0% train, rema	ainder test		
ult list (right-click for options)	=== Classifier mode	l (full training)	set) ===		
7:06:47 - haves NaiveBaves	Naive Bayes Classif	fier			
7:07:05 - trees.REPTree		C 1			
7:07:14 - functions.LibSVM	Attribute	Iris-setosa Iris-	versicolor Iris	virginica	
		(0.33)	(0.33)	(0.33)	
	sepallength				
	std. dev.	4.9913 0.355	5.9379 0.5042	6.5795 0.6353	
	weight sum	50	50	50	
	president	011005	012005	011003	
	sepalwidth	2 4015	0.000	2.0020	
us					
<					Log 📣
reprocess Classify Cluster Associ ssifier Choose LibSVM -S 0 -K 2 -D 3 -G 0.0	(ate Select attributes Visuali	ize 	\\Program Files\\\Ve	a-3-8-4" - seed 1	
t options	Classifier output				
Use training set					
) Supplied test set Set	sepalwidth				
Cross-validation Folds 10	mean atd day	3.4015	2.7687	2.9629	
Percentage split % 66	weight sum	50	50	50	
More ontions	precision	0.1091	0.1091	0.1091	
more options	petallength	1.4004	4 2452	5 5514	
m) class	std. dev.	0.1782	4.2452	0.5529	
	weight sum	50 0.1405	50 0.1405	50 0.1405	
Start Stop					
ult list (right-click for options)	petalwidth mean	0.2743	1.3097	2.0343	
7:06:47 - bayes.NaiveBayes	std. dev.	0.1096	0.1915	0.2646	
07:05 - troop REPTroo	precision	50	30	30	
107:14 - functions LibSVM		0.1143	0.1143	0.1143	
207:14 - functions.LibSVM		0.1143	0.1143	0.1143	
7:07:14 - functions.LibSVM	Time prime as by 199	0.1143	0.1143	0.1143	
7:07:14 - functions.LibSVM	Time taken to build	0.1143 Model: 0 seconds	0.1143	0.1143	
17:07:14 - functions LibSVM	Time taken to build	0.1143 i model: 0 seconds xest split ===	0.1143	0.1143	
17:07:14 - functions LibSVM	Time taken to build === Evaluation on t Time taken to test :	0.1143 1 model: 0 seconds :est split === model on test spl:	0.1143 it: 0.01 seconds	0.1143	
7:07:14 - functions.LibSVM	Time taken to build === Evaluation on t Time taken to test : === Summary ===	0.1143 i model: 0 seconds est split === model on test spl:	0.1143 it: 0.01 seconds	0.1143	
7.07.14 - functions LibSVM	Time taken to build Evaluation on t Time taken to test : Summary	0.1143 i model: 0 seconds :est split === model on test spl:	0.1143 it: 0.01 seconds	0.1143	
tus	Time taken to build Evaluation on t Time taken to test : Summary	0.1143 i model: 0 seconds :est split === model on test spl:	0.1143 it: 0.01 seconds	0.1143	
tus	Time taken to build Evaluation on t Time taken to test : Summary	0.1143 I model: 0 seconds est split === model on test spl:	0.1143 it: 0.01 seconds	0.1143	Log

Naïve Bayes summary

🥝 Weka Explorer							-	٥	×
Preprocess Classify Cluster Associate	Select attributes Visualize								
Classifier									
	0.0 N.0.6 M.40.0 C.1.0 E.0.001 B.0.1 mod	ol "C:\\Brogrom Filoo\\\#(olv	3.8.45 cood 1						
CHOOSE LIDSOM -5 0 -K 2 -D 3 -G 0.0 -R	0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -M00	er C. (Program Files (Wieka	+3-8-4 -Seeu I						
Test options	Classifier output								
 Use training set 	Time taken to test model on test	apiic. 0.01 aeconua							
O Supplied test set Set	=== Summary ===								
O Cross-validation Folds 10	Correctly Classified Instances	48	94.1176 %						
Percentage split % 66	Incorrectly Classified Instances	3	5.8824 %						
	Kappa statistic	0.9113							
More options	Root mean squared error	0.1722							
	Relative absolute error	10.0365 %							
(Nom) class	Root relative squared error Total Number of Instances	36.4196 %							
Start Stop		01							
Result list (right-click for options)	=== Detailed Accuracy by class =								
	TP Rate FP Rat	e Precision Recall	F-Measure MCC	ROC Area	PRC Area	Class			
17:06:47 - bayes.NaiveBayes	1.000 0.000	1.000 1.000	1.000 1.000	1.000	1.000	Iris-setosa Iris-versicolor			
17:07:14 - functions LibSVM	0.882 0.029	0.938 0.882	0.909 0.867	0.988	0.980	Iris-virginica			
17.07.14 - Iuliciolis.Elbovii	Weighted Avg. 0.941 0.033	0.942 0.941	0.941 0.909	0.992	0.986				
	Confusion Matrix								
	15 0 0 a = Iris-setosa								
	0 18 1 b = Iris-versicolor								
	0 2 15 c = Iris-virginica								
									-
Status									
ок							Log	~	× 0
🖶 🛱 含 👼 💼 f	🚖 🔺 🔇 🥥 🗷 🛯						^ %ED 🦟 Φ)) ENG 🕺	:09 PM 26/2020	-

Applying Decision Tree

We are going to select classifier and choose REPTree (which is also Decision Tree).



Decision Tree summary:

	Webs	Evn	lorer
-	TT CKU	LAP	loici

Preprocess Classify Cluster Associate	Select attributes Visualize					
Classifier						
Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.	0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model*	'C:\\Program Files\\Weka	-3-8-4" -seed 1			
Test options	Classifier output					
O Use training set	Time taken to test model on test ap	piit. o seconus				A
Supplied test set Set	=== Summary ===					
Cross-validation Folds 10 Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error	47 4 0.8824 0.058 0.1963 13.0214 %	92.1569 % 7.8431 %			
(Nom) class	Root relative squared error Total Number of Instances	41.5318 % 51				
Start Stop	=== Detailed Accuracy By Class ===					
Result list (right-click for options) 17:06:47 - bayes. NaiveBayes 17:07:05 - trees REPTree 17:07:14 - functions. LibSVM	TP Rate FP Rate 1.000 0.000 0.789 0.000 1.000 0.118 Weighted Avg. 0.922 0.039 === Confusion Matrix === a b c < classified as 15 0 0 a = Iris-setosa 0 15 4 b = Iris-versicolor	Precision Recall 1.000 1.000 1.000 0.789 0.810 1.000 0.937 0.922	F-Measure MCC 1.000 1.000 0.882 0.888 0.895 0.845 0.921 0.888	ROC Area PRC Area 1.000 1.000 0.972 0.950 0.971 0.895 0.980 0.946	Class Iris-setosa Iris-versicolor Iris-virginica	
Siatus OK	0 0 17 c = Iris-virginica					

- o ×



٥ 🕝 Weka Explorer _ × Preprocess Classify Cluster Associate Select attributes Visualize Classifier Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4" -seed 1 Test options Classifier output 🔘 Use training set === Run information === ○ Supplied test set Scheme: Relation: Instances: Attributes: weka.classifiers.functions.Lib5VM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\\Program Files\\Weka-3-O Cross-validation Folds 10 iris 150 Percentage split % 66 5 Attributes: 5 sepallength sepalvidth petallength petalwidth class Test mode: split 66.0% train, remainder test More options... (Nom) class • Start Stop === Classifier model (full training set) === Result list (right-click for options) 17:06:47 - bayes.NaiveBayes LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM) 17:07:05 - trees.REPTree 17:07:14 - functions.LibSVM Time taken to build model: 0.01 seconds === Evaluation on test split === Time taken to test model on test split: 0.01 seconds === Summary === Correctly Classified Instances Incorrectly Classified Instances 51 0 100 0 40 ale 7 1 Status Log x0 ок 🕂 🛱 🧲 🏦 💼 🔺 🔇 🧿 W 垦

SVM summary:

Weka Explorer												- 0) ×
Preprocess Classify Cluster Associate Sel	elect attributes 🏹 Vis	ualize											
Classifier													
Choose LibSVM-S0-K2-D3-G00-R00-N	105-M400-C10	E 0 001 -P	0.1 -model	"C:))Program I	Files)WA(eks	.3.8.4" .cood	1						_
				one rogram.									
Test options Clas	ssifier output												
◯ Use training set													
O Supplied test set Set	== Summary ===												
	orrectly Classi:	fied Inst	ances	51		100	\$						
Cross-validation Folds 10	ncorrectly Class	sified In	stances	0		0	ę						
 Percentage split % 66 Ka 	appa statistic			1									
Men	lean absolute er:	ror		0									
More options	oot mean square	a error		0									
Re	oot relative so	uared err	or	0	8								
(Nom) class 🔹 To	otal Number of	Instances		51									
Start Stop	== Detailed Acc	uracy By	Class ===										
Result list (right-click for options)		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class			
		1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-setosa			
17:06:47 - bayes.NaiveBayes		1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-versicolor			
17:07:05 - trees.REPTree		1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-virginica			
17:07:14 - functions.LibSVM	eighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000				
	== Confusion Ma	trix ===											
	a b c (classifia	d as										
	15 0 0 a =	Iris-set	osa										
	0190 b =	Iris-ver	sicolor										
	0 0 17 c =	Iris-vir	ginica										
													7.
status													
ок												Log	
				_									-
🕂 🛱 🧲 📄 🔒	🔺 🕓	O	R 👯								へ 幅 <i>信</i> (10) EM	IG 1/26/20	M 🖣

Comparing the results:

Naïve Bayes:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
0.900	0.947	0.923	0.876
0.938	0.882	0.909	0.867
0.942	0.941	0.941	0.909

Decision tree:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
1.000	0.789	0.882	0.838
0.810	1.000	0.895	0.845
0.937	0.922	0.921	0.888

NVM:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
0.959	0.940	0.949	0.925
0.941	0.960	0.950	0.925
0.967	0.967	0.967	0.950