Ahmad Ali

ID: 13473 SE-07

Data Mining

Submitted To: Sir Zain Shukat

Q1:

Ans:

Classification

- The discovery of model or functions where the mapping of objects is done into predefined classes.
- Discrete values
- Decision tree, logistic regression, etc.

• Unordered

• Measuring accuracy

Regression

- A devised model in which the mapping of objects is done into values.
- Continuous values
- Regression tree (Random forest), Linear regression, etc.
- Ordered
- Measurement of root mean square error

Key Differences Between Classification and Regression

- 1. The Classification process models a function through which the data is predicted in discrete class labels. On the other hand, regression is the process of creating a model which predict continuous quantity.
- 2. The classification algorithms involve decision tree, logistic regression, etc. In contrast, regression tree (e.g. Random forest) and linear regression are the examples of regression algorithms.
- 3. Classification predicts unordered data while regression predicts ordered data.
- 4. Regression can be evaluated using root mean square error. On the contrary, classification is evaluated by measuring accuracy.

Classification example:

Suppose from your past data (train data) you come to know that your best friend likes the above movies. Now one new movie (test data) released. Hopefully, you want to know your best friend like it or not. If you strongly confirmed about the chances of your friend like the move. You can take your friend to a movie this weekend.

If you clearly observe the problem it is just whether your friend like or not. Finding a solution to this type of problem is called as classification. This is because we are classifying the things to their belongings (yes or no, like or dislike). Keep in mind here we are forecasting target class (classification) and the other thing this classification belongs to supervised learning. This is because you are learning this from your train data.

In this case, the problem is a binary classification in which we have to predict whether output belongs to class 1 or class 2 (class 1: yes, class 2: no). As we have discussed earlier, we can use classification for predicting more classes too. Like (color prediction: red,green,blue,yellow,orange)

Regression example:

Suppose from your past data (train data) you come to know that your best friend likes the above movies. You also know how many times each particular movie seen by your friend. Now one new movie (test data) released. Now you are going to find how many times this newly released movie will your friend watch. It could be 5 times, 6 times, 10 times etc...

If you clearly observe the problem is about finding the count, sometimes we can say this as predicting the value. Keep in mind, here we are forecasting a value (prediction) and the other thing this prediction also belongs to supervised learning. This is because you are learning this from you train data.

Q2:

Ans:

🖉 Weka Explorer								- C	x נ	
Preprocess Classify Cluster Associate	Select attributes	Visualize								
Classifie										
Choose NaiveBayes										
Test option	Classifier outpu									
Use training set Supplied test set Cross-validation Endds	=== Run information === Scheme: weka.classifiers.bayes.NaiveBayes									
Percentage split % 66	Instances:	8								
	Attributes:	4 Xu : Toat l								
More options		Test2								
		Final								
(Nom) Grade	Test mode:	evaluate on	training	data						
Start Stop	=== Classifia	er model (ful	1 trainin	a set) ==	-					
Result list (right-click for option 23:18:31 - bayes.NaiveBayes	Naive Bayes Classifier									
	2000002000000	Class		-			-	_		
	Attribute	D (0.2)	C (0.13)	B- (0.13)	A- (0.13)	C-	F (0.13)	B+ (0.13)		
		(0.12)	(0120)	(0.120)	(0.120)	(0120)	(0.120)	(0120)		
	Testl									
	mean	101	57.7143	28.8571	86.5714	86.5714	2 4049	28.8571		
	weight sum	2.4048	2.4048	2.4048	2.4040	2.4048	2.4048	2.4040		
	precision	14.4286	14.4286	14.4286	14.4286	14.4286	14.4286	14.4286		
	Test2									
	mean	85.25	46.5	46.5	77.5	93	0	46.5		
	std. dev.	7.75	2.5833	2.5833	2.5833	2.5833	2.5833	2.5833		
	weight sum	2	1	1	1	1	1	1		
	precision	15.5	15.5	15.5	15.5	15.5	15.5	15.5		
	Final									
	mean	48.5	44	47	45	46	43	50		
	std. dev.	0.5	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667		
	weight sum	2	1	1	1	1	1	1		
	precision	1	1	1	1	1	1	1		
							,			
Status										
ок								Log	🐨 ×0	

Weka Explor	rer									<u>100</u> 7		×
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize							
lassifie												
Choose	NaiveBa	/es										
est option				Classifier outpu								
🔿 Use train	ing set			Time taken t	- build model	. 0	ada					
Supplied	test set	Set.		Time taken t	o build model	. o secor	lus					
0.0				=== Evaluati	on on trainin	g set ===	-					
O Cross-va	lidation i	-olds /		Time taken t	a test model	on train	ng data. 0	seconds				1
O Percenta	ge split	% 66	ò	Time caken c	o cest moder	on train.	ing data: 0	seconds				1
[More optic	ons		=== Summary								н
				Compart 1			0		100			
				Incorrectly CI	assified inst Classified In	stances	8		001	95 95		11
Nom) Grade				Kappa statis	tic		1			0		11
				Mean absolut	e error		0					
Start		Sto	p	Root mean sq	uared error		0					
esult list (rio	ht-click f	or optior		Relative abs	o <mark>lute error</mark>		0.00	01 %				
				Root relativ	e squared err	or	0.00	01 %				
23:18:31 - b	aves.Naive	Bayes		Total Number	of Instances		8					
				=== Detailed	Accuracy By	Class ===						1
					TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Are	a
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
					1.000	0.000	1.000	1.000	1.000	1.000	1.000	
				Weighted Avg	1.000	0.000	1.000	1.000	1.000	1.000	1.000	
				=== Confusio	n Matrix ===							
				abcdei	g <-= clas	silled a:	3					
				200000	0 a = D							
				0 0 1 0 0 0	$0 \mid c = B_{-}$							
				0 0 0 1 0 0	0 I d = 2-							
				0 0 0 0 1 0	0 e = C-							
				0 0 0 0 0 1	0 I f = F							
				000000	1 q = B+							
				4)		
										_		-
tatus												
												~
OK										20		r

Q3:

Ans:

- First download any dataset from any source you like. (make sure it is in .arff extension)
- Open Weka software.
- Click on the explorer button.
- Then click on open file button on top left corner.
- Browse to file you've just downloaded
- Add filters if you want.
- Then click on classify button on top.
- Apply various classifier algorithms (naive Bayes, J48 etc.).

Weka Explorer									- 0	1 X
Preprocess Classify Cluster Associate	Select attributes Vi	sualize								
Classifie										
J48 -C 0.25 -M 2										
Test option:	Classifier outpu									
	vio pruneu cree									
O Use training set		1000								A
O Supplied test set Set										
Cross-validation Folds 10	outlook = sunny	75. 100 1	(2.0)							
	humidity > 7	5: no (3.	.0)							
O Percentage spilt % 66	outlook = overca	st: yes ((4.0)							
More options	outlook = rainy									
	windy = TRUE	: no (2.0))							
	windy = FALS	E: yes (3	3.0)							
(Nom) play	Number of Leaves									
	Number of heaves		,							
Start Stop	Size of the tree	: 8	3							
Result list (right-click for option										
	Time to be the back			terrenerative.						
01:16:01 - trees.J48	lime taken to bu	illa model	.: U.UI se	conas						
	=== Stratified o	ross-vali	dation ==	=						
	=== Summary ===									
	Correctly Classi	fied Inst	ances	9						
	Incorrectly Clas	sified In	stances	5						
	Kappa statistic			0.18						
	Root mean square	d error		0.48	18					
	Relative absolut	e error		60 %						
	Root relative so	puared err	or	97.65						
	Total Number of	Instances	3	14						
		-	~							
	=== Detailed Acc	uracy By	Class ===							
		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Clas
		0.778	0.600	0.700	0.778	0.737	0.189	0.789	0.847	yes
		0.400	0.222	0.500	0.400	0.444	0.189	0.789	0.738	no
	Weighted Avg.	0.643	0.465	0.629	0.643	0.632	0.189	0.789	0.808	
	=== Confusion Ma	trix ===								
	CONTROLOG IN									
	a b < class	ified as								
	72 a=yes									
	3 2 b = no									
										~
	-								;	
[]										
Status							_			
ок									Log	×0

😋 Weka Explo	orer								- <u></u> -		×
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize						
Classifie	_			•							
)										-
Choose	NaiveBa	yes									
Test option				Classifier outpu							
O Use trair	ning set										
 O Supplier 	l to at ant	0.4		=== Run infor	rmation ===						5
O Supplied	i lest set	Ser		Scheme:	weka.clas	sifiers.	ayes.NaiveBayes				
Cross-va	alidation	Folds 1	0	Relation:	weather						
O Percenta	age split	% 6	6	Instances:	14						
				Attributes:	5						
	More optio	ons			temperatu	re					
					humidity						
(Nom) play					windy						
(, [)					play						
Start		St	op	Test mode:	10-fold c	ross-val:	dation.				
Rocult liet (ri	aht click f	or ontiou		=== Classifie	er model (f	ull train	ning set) ===				
ives une not (in	gint-click i	or option					5 N				
01:16:01 - t	rees.J48			Naive Bayes C	Classifier						
01:20:18 - b	oayes.Naive	eBayes									
				Attribute	Class						
				ACCLIDACE	(0.63)	(0.38)					
				outlook							
				sunny	3.0	4.0					
				overcast	5.0	1.0					
				ftotall	4.0	3.0					
				[cotar]	12.0	0.0					- 1
				temperature							- 1
				mean	72.9697	74.8364					- 1
				std. dev.	5.2304	7.384					- 8
				weight sum	1 0001	1 0001					- 8
				precision	1.5051	1.5051					- 8
				humidity							
				mean	78.8395	86.1111					- 1
				std. dev.	9.8023	9.2424					- 1
				weight sum	9	5					- 1
				precision	3.4444	3.4444					- 1
				windy							
				TRUE	4.0	4.0					
				FALSE	7.0	3.0					
				[total]	11.0	7.0					v
				-							
Status				· · · · · · · · · · · · · · · · · · ·							
Status											
OK									LOS	۳ A	er x

Weka Explorer

	V			
_	~			

Preprocess	Classify	Cluster	Associate	Select attributes	Visualize										
Classifie															
	i													_	
Choose	NaiveBay	res													
Test option				Classifier outpu											
0														-	
O Use train	ing set			weight sum	9	5								A	
O Supplied	test set	Set.		precision	1.9091	1.9091									
Cross-va	lidation F	olds 10		humidity											
O Percenta	ige split	% 66		mean	78.8395	86.1111									
-				std. dev.	9.8023	9.2424									
	More optio	ns		weight sum	2 4444	3 4444									
]	precision	5.1111	5.1111									
[windy											
(Nom) play				TRUE	4.0	4.0									
C				FALSE	7.0	3.0									
Start		Sto	P	[total]	11.0	7.0									
Result list (rig	ht-click fo	or option													
01:49:44 - b	ayes.Naive	Bayes													
				Time taken to	build mode	el: 0 seco	nds								
				=== Summary =	==	riuacion -									
				Correctly Cla	ssified In:	stances	9		64.2857	alo					
				Incorrectly C	lassified :	Instances	5		35.7143	35.7143 %					
				Kappa statist	ic		0.10	0.1026							
				Mean absolute	Mean absolute error			549							
				Root mean squ	ared error		0.54	13							
				Relative abso	Relative absolute error				97.6254 %						
				Root relative	squared en	rror	110.051 %								
				Total Number	of instance	es	14								
				=== Detailed	Accuracy B	y Class ==	-								
					TP Rate	e FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class		
					0.889	0.800	0.667	0.889	0.762	0.122	0.444	0.633	yes		
					0.200	0.111	0.500	0.200	0.286	0.122	0.444	0.397	no		
				Weighted Avg.	0.643	0.554	0.607	0.643	0.592	0.122	0.444	0.548			
				=== Confusion	Matrix ===	=									
🦻 Stud	ent Inform	ation Cent	er	a b < cl	assified a:	5									
And Annual Annua	And a second sec			81 a = ye	s										
O-104 Patent	alard how o	97 - anar - arren - hel	- 4	4 1 b = no											
Upload Ansigne	rents														
														V	
Charles and an and a second	nd, de El providel, Prispetialed, Vasaanned Jr.D. Pari en Despolaarring-boldeline et Besoarb Vasaanne Despolaarring-boldeline et Besoarb	nanta geografi, «Man a Negel generi, Petrovali na tec	- 1												
Stati										_					
OK												Log	100	× x0	
(And a state	-														