Fall 2020: CSCI 4/5588 Programming Assignment #2

Name: Tharani Maaneeivaannan and ID:2575198

Question-1: Describe each classifier in 5 to 10 sentences and provide an appropriate reference(s).

Classifier-1 SMO/SVM (Support Vector Machine)

SVM is a hyperplane that separates a set of positive examples from a set of negative examples with maximum margin. In the linear case, the margin is defined by the distance of the hyperplane to the nearest of the positive and negative examples.

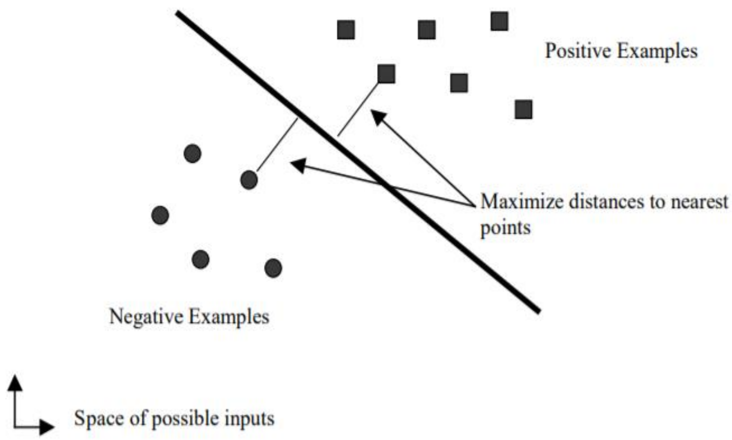

Figure 1 Linear Support Vector Machine

The new SVM learning algorithm is called Sequential Minimal Optimization (or SMO). Instead of previous SVM learning algorithms that use numerical quadratic programming (QP) as an inner loop, SMO uses an analytic QP step. Sequential Minimal Optimization (SMO) is a simple algorithm that can quickly solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps at all.

Classifier-2 Lazy IBK

IBK algorithm implements the k-nearest neighbor algorithm. It works by storing the entire training dataset and querying it to locate the k most similar training patterns when making a prediction. As such, there is no model other than the raw training dataset and the only computation performed is the querying of the training dataset when a prediction is requested. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. Instance-based learning generates classification predictions using only specific instances. Instance-based learning algorithms do not maintain a set of abstractions derived from specific instances.

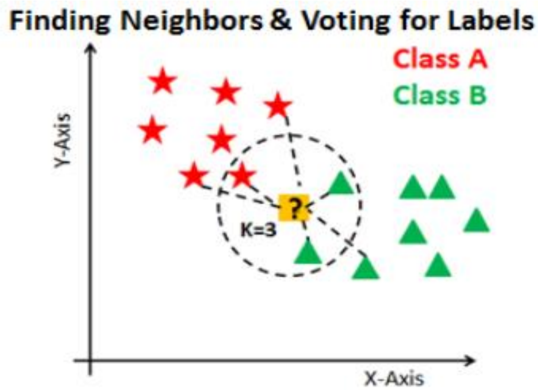

Figure 2 K-NN algorithm

Classifier-3 Naïve Bayes

It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, a Naive Bayes classifier would consider all of these properties to independently contribute to the probability. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is also known to outperform even highly sophisticated classification methods. Naive Bayes is a supervised learning algorithm for classification which will find the class of observation (data point) given the values of features. Naive Bayes classifier calculates the probability of a class given a set of feature values. The algorithm needs to store probability distributions of features for each class independently. The type of distributions depend on the characteristics of features:

For binary features (Y/N, True/False, 0/1): Bernoulli distribution

For discrete features (i.e. word counts): Multinomial distribution

For continuous features: Gaussian (Normal) distribution

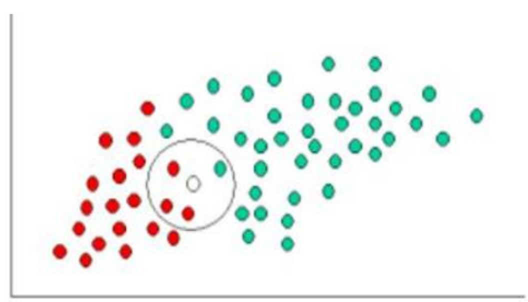

Figure 3 Naive Bayes classifier

Classifier-4 Trees J48

A decision tree be a flowchart-like tree structure, where each internal node represents a test happening an attribute, each branch represents an ending of the test, class label is represented by each leaf node or terminal node. Quinlan's C4.5 algorithm actualizes J48 to create a trimmed C4.5 decision tree. The information is split into minor subsets. The minor subsets are returned by the algorithm. The split strategies stop if a subset has a place with a similar class in all the instances. J48 decision tree can deal with particular characteristics, lost or missing attribute estimations of the data and varying attribute cost.

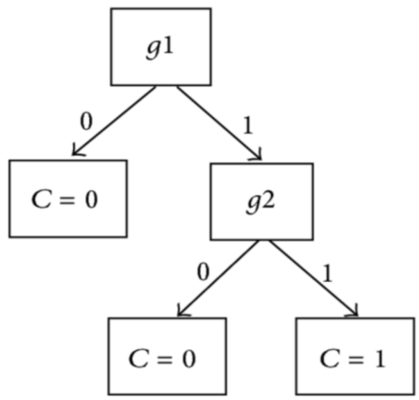

Figure 4 J-48 tree

Classifier-5 Multilayer Perceptron

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not a linearly separable. The perceptron is an algorithm for supervised learning of binary classifiers. A binary classifier is a function decide whether input belongs to some specific class based on a linear predictor function combining a set of weights with the feature vector.

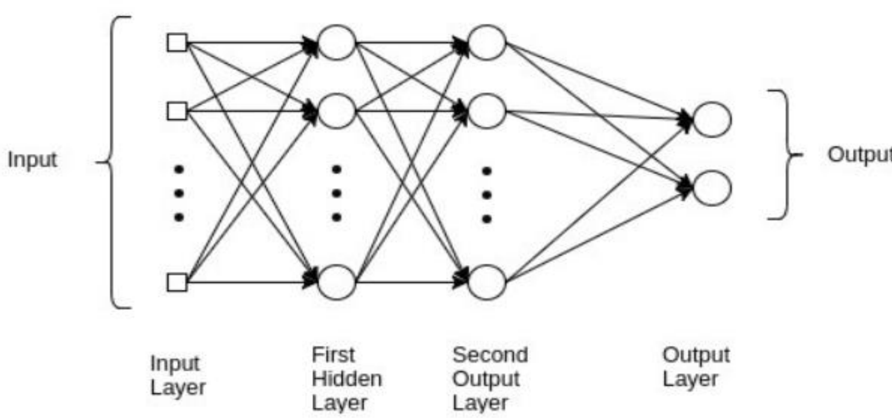

Figure 5 Multilayer Perceptron

References:

1. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-14.pdf
2. https://link.springer.com/article/10.1023/A:1022689900470
3. https://towardsdatascience.com/naive-bayes-classifier-explained-50f9723571ed

4. https://arxiv.org/ftp/arxiv/papers/1302/1302.4964.pdf
5. https://link.springer.com/article/10.1007/BF00993309
6. https://www.researchgate.net/publication/266396438_A_Gentle_Introduction_to_Backpropagation

Question-2: Describe the parameters/hyper-parameters that you have chosen to train the classifiers.

SMO

-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K

```
Kernel used:
  Linear Kernel: K(x,y) = <x,y>

Classifier for classes: 0, 1

BinarySMO

Machine linear: showing attribute weights, not support vectors.

        0.6738 * (normalized) Sbp
 +      1.9215 * (normalized) Tobacco
 +      1.9395 * (normalized) Ldl
 +      0.4514 * (normalized) Adiposity
 +      0.8898 * (normalized) Famhist=Present
 +      1.8288 * (normalized) Typea
 +     -1.3137 * (normalized) Obesity
 +     -0.1412 * (normalized) Alcohol
 +      1.3164 * (normalized) Age
 -      3.6287

Number of kernel evaluations: 18856 (68.705% cached)
```

Lazy IBK

weka.classifiers.lazy.IBk -K 1 -W 0 -A

J-48

weka.classifiers.trees.J48 -C 0.25 -M 2

Naïve Bayes

weka.classifiers.bayes.NaiveBayes

Multilayer perceptron

weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Question-3: Define and describe the following terms (provide an appropriate reference(s)) for measuring the performances of a classifier:

True Positive (TP) rate- refers to a situation where predicted as positive are set of positive examples.

False Positive (FP) rate- refers to a situation where predicted as positive but they are set of negative examples.

Precision- Precision is also known as the positive predictive value. Precision is the number of correct positive results divided by the number of positive results predicted by the classifier. Precision is defined as the accuracy of the judgment. $Precision = TP/(TP+FP)$

Recall- Recall is also known as the true positive rate, sensitivity, probability of detection, hit rate. Recall is the ratio of predicted positive data with respect to all positive data. $\text{Recall} = TP/(TP+FN)$

F-Measure- F-measure combines both precision and recall which is calculated as percentages and combined as harmonic mean to assign a single number. This is also known as F_1 score or balanced F-score. F- measure=2 $(precision \times recall)/(precision+recall) = (2\,TP)/(2TP+FP+FN)$

Receiver Operating Characteristic (ROC) Area- ROC curve is a graph showing the TPR (True Positive Rate) and FPR (False Positive Rate) at thresholds. AUC is the area under the curve of the ROC. AUC close to or equal to 1.0 indicates the best performance. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s.

Confusion Matrix- A Confusion matrix is an N x N matrix used for evaluating the accuracy of a classification model, where N is the number of target classes. The confusion matrix is as follows:

		Actual class	
		P	**N**
Predicted class	**P**	**TP**	FP
	N	FN	**TN**

where: P = Positive; N = Negative; TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

Question-4: Describe the performance of the classifiers in terms of

Correctly Classified Instances

5-FCV

SMO

Correctly Classified Instances	323	69.9134 %
Incorrectly Classified Instances	139	30.0866 %
Kappa statistic	0.3143	
Mean absolute error	0.3009	
Root mean squared error	0.5485	
Relative absolute error	66.4131 %	
Root relative squared error	115.2823 %	
Total Number of Instances	462	

Lazy IBk

Correctly Classified Instances	292	63.2035 %
Incorrectly Classified Instances	170	36.7965 %
Kappa statistic	0.1727	
Mean absolute error	0.3687	
Root mean squared error	0.605	
Relative absolute error	81.3815 %	
Root relative squared error	127.1487 %	
Total Number of Instances	462	

NaiveBayes

Correctly Classified Instances	325	70.3463 %
Incorrectly Classified Instances	137	29.6537 %
Kappa statistic	0.3592	
Mean absolute error	0.3271	
Root mean squared error	0.4775	
Relative absolute error	72.2014 %	
Root relative squared error	100.3666 %	
Total Number of Instances	462	

J48

Correctly Classified Instances	308	66.6667 %
Incorrectly Classified Instances	154	33.3333 %
Kappa statistic	0.2322	
Mean absolute error	0.3896	
Root mean squared error	0.4977	
Relative absolute error	86.0003 %	
Root relative squared error	104.6002 %	
Total Number of Instances	462	

Multilayer Perceptron

Correctly Classified Instances	307	66.4502 %
Incorrectly Classified Instances	155	33.5498 %
Kappa statistic	0.2212	
Mean absolute error	0.3733	
Root mean squared error	0.4994	
Relative absolute error	82.4108 %	

Root relative squared error 104.9536 %
Total Number of Instances 462

	SMO	Lazy IBK	Naïve Bayes	J48	Multilayer Perceptron
Correctly Classified Instances	323	292	325	308	307
Incorrectly Classified Instances	139	170	137	154	155
Kappa statistic	0.3143	0.1727	0.3592	0.2322	0.2212
Mean absolute error	0.3009	0.3687	0.3271	0.3896	0.3733
Root mean squared error	0.5485	0.605	0.4775	0.4977	0.4994
Relative absolute error	66.4131 %	81.3815 %	72.2014 %	86.0003 %	82.4108 %
Root relative squared error	115.2823 %	127.1487 %	100.3666 %	104.6002 %	104.9536 %
Total Number of Instances	462	462	462	462	462

10 FCV

SMO

Correctly Classified Instances	328	70.9957 %
Incorrectly Classified Instances	134	29.0043 %
Kappa statistic	0.3299	
Mean absolute error	0.29	
Root mean squared error	0.5386	
Relative absolute error	64.028 %	
Root relative squared error	113.1898 %	
Total Number of Instances	462	

Lazy IBk

Correctly Classified Instances	292	63.2035 %
Incorrectly Classified Instances	170	36.7965 %
Kappa statistic	0.1652	
Mean absolute error	0.3686	
Root mean squared error	0.6052	
Relative absolute error	81.3691 %	
Root relative squared error	127.1864 %	
Total Number of Instances	462	

NaiveBayes

Correctly Classified Instances	331	71.645 %
Incorrectly Classified Instances	131	28.355 %
Kappa statistic	0.3855	
Mean absolute error	0.3238	
Root mean squared error	0.4725	
Relative absolute error	71.4816 %	
Root relative squared error	99.3063 %	
Total Number of Instances	462	

J48

Correctly Classified Instances	327	70.7792 %
Incorrectly Classified Instances	135	29.2208 %
Kappa statistic	0.328	
Mean absolute error	0.3689	
Root mean squared error	0.4733	
Relative absolute error	81.4419 %	
Root relative squared error	99.4841 %	
Total Number of Instances	462	

Multilayer Perceptron

Correctly Classified Instances	316	68.3983 %
Incorrectly Classified Instances	146	31.6017 %
Kappa statistic	0.2765	
Mean absolute error	0.3516	
Root mean squared error	0.4749	
Relative absolute error	77.6125 %	
Root relative squared error	99.8205 %	
Total Number of Instances	462	

	SMO	Lazy IBK	Naïve Bayes	J48	Multilayer Perceptron
Correctly Classified Instances	328	292	331	327	316
Incorrectly Classified Instances	134	170	131	135	146
Kappa statistic	0.3299	0.1652	0.3855	0.328	0.2765
Mean absolute error	0.29	0.3686	0.3238	0.3689	0.3516
Root mean squared error	0.5386	0.6052	0.4725	0.4733	0.4749
Relative absolute error	64.028 %	81.3691 %	71.4816 %	81.4419 %	77.6125 %
Root relative squared error	113.1898 %	127.1864 %	99.3063 %	99.4841 %	99.8205 %
Total Number of Instances	462	462	462	462	462

SMO

5-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.805	0.500	0.752	0.805	0.778	0.316	0.652	0.733	0
	0.500	0.195	0.576	0.500	0.535	0.316	0.652	0.461	1
Weighted avg.	0.699	0.394	0.691	0.699	0.694	0.316	0.652	0.639	

10-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.828	0.513	0.753	0.828	0.789	0.334	0.658	0.736	0
	0.488	0.172	0.600	0.488	0.538	0.334	0.658	0.470	1
Weighted avg.	0.710	0.395	0.700	0.710	0.702	0.334	0.658	0.644	

Lazy IBK

5-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.738	0.569	0.710	0.738	0.724	0.173	0.590	0.701	0
	0.431	0.262	0.466	0.431	0.448	0.173	0.590	0.412	1

Weighted avg.	0.632	0.462	0.626	0.632	0.628	0.173	0.590	0.601	

10-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.748	0.588	0.706	0.748	0.727	0.166	0.587	0.698	0
	0.413	0.252	0.465	0.413	0.437	0.166	0.587	0.409	1
Weighted avg.	0.632	0.471	0.623	0.632	0.626	0.166	0.587	0.598	

Naïve Bayes

5-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.748	0.381	0.787	0.748	0.767	0.360	0.745	0.842	0
	0.619	0.252	0.566	0.619	0.591	0.360	0.745	0.567	1
Weighted avg.	0.703	0.336	0.711	0.703	0.706	0.360	0.745	0.747	

10-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.762	0.369	0.796	0.762	0.778	0.386	0.749	0.843	0
	0.631	0.238	0.584	0.631	0.607	0.386	0.749	0.580	1
Weighted avg.	0.716	0.324	0.722	0.716	0.719	0.386	0.749	0.752	

J-48

5-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.791	0.569	0.724	0.791	0.756	0.234	0.647	0.742	0
	0.431	0.209	0.523	0.431	0.473	0.234	0.647	0.450	1
Weighted avg.	0.667	0.444	0.654	0.667	0.658	0.234	0.647	0.64	

10-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.821	0.506	0.754	0.821	0.786	0.331	0.667	0.754	0
	0.494	0.179	0.594	0.494	0.539	0.331	0.667	0.481	1
Weighted avg.	0.708	0.393	0.698	0.708	0.701	0.331	0.667	0.660	

Multilayer Perceptron

5-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.798	0.588	0.719	0.798	0.757	0.224	0.657	0.787	0
	0.413	0.202	0.520	0.413	0.460	0.224	0.657	0.488	1
Weighted avg.	0.665	0.453	0.650	0.665	0.654	0.224	0.657	0.684	

10-FCV

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.798	0.531	0.739	0.798	0.768	0.278	0.706	0.808	0
	0.469	0.202	0.551	0.469	0.507	0.278	0.706	0.546	1
Weighted avg.	0.684	0.417	0.674	0.684	0.677	0.278	0.706	0.717	

Confusion Matrix

SMO

5-FCV

```
  a   b   <-- classified as
243  59 |   a = 0
 80  80 |   b = 1
```

10-FCV

```
  a   b   <-- classified as
250  52 |   a = 0
 82  78 |   b = 1
```

Lazy IBK

5-FCV

```
  a   b   <-- classified as
223  79 |   a = 0
 91  69 |   b = 1
```

10-FCV

```
   a   b   <-- classified as
 226  76 |   a = 0
  94  66 |   b = 1
```

Naïve Bayes

5-FCV

```
   a   b   <-- classified as
 226  76 |   a = 0
  61  99 |   b = 1
```

10-FCV

```
   a   b   <-- classified as
 230  72 |   a = 0
  59 101 |   b = 1
```

J-48

5-FCV

```
   a   b   <-- classified as
 239  63 |   a = 0
  91  69 |   b = 1
```

10-FCV

```
   a   b   <-- classified as
 248  54 |   a = 0
  81  79 |   b = 1
```

Multilayer perceptron

5-FCV

```
   a   b   <-- classified as
 241  61 |   a = 0
  94  66 |   b = 1
```

10-FCV

```
   a   b   <-- classified as
 241  61 |   a = 0
  85  75 |   b = 1
```