Homework 4

1.)
a.) 2 necessary conditions for an ensemble classifier to perform better than a base classifier:
i.) The base classifiers should be independent of one another.
ii.) The base classifier should do better than a classifier that performs random guessing.

b.) Ways to ensure total independence among classifiers:
-Split the dataset into uncorrelated parts and train classifiers on them.
-Using multiple types of unstable classifiers.

c.) Unstable classifiers:
Unstable classifiers are base classifiers that are sensitive to minor perturbations in the training set. Decision trees, rule-based classifier and artificial neural networks are unstable classifiers.

d.) Procedure for modified ensemble method:

-Each D_i is a bootstrap sample having only a subset of training features so basically we need to perform bagging.
-Use unstable classifiers as base classifiers: for eg. Decision trees.
-Perform accuracy weighted prediction to perform class label: If the accuracy of classifier C_i is a_i, we assign a weight to it while voting w_i proportional to a_i.

1: Let D denote the original training data, k denote the number of base classifiers,
And T be the test data
2: for i=1 to k do
3: Create training set, D_i from D using bagging approach
4: Build an unstable base classifier C_i from D_i.
5: assign w_i weight based on accuracy of C_i
6: end for
7: for each test record x belonging to T do
8: C*(x) = Vote (w1*C1(x),w2*C2(x),.....wk*C2(k))
9: end for

2.)

a.) i.) Contingency Table:

	Predicted Class		
Actual Class		Class = +	Class = -
	Class = +	1 (a)	1 (b)
	Class= -	2 (c)	10 (d)

ii.) Precision(p) = a/(a+c) = 1/(1+2) = 1/3 = 0.333

iii.) Recall (r) = a/(a+b) = 1/(1+1) = ½ = 0.5

iv.) F-measure = 2rp/(r+p) = 2x0.5x0.333/(0.5+0.333) = 0.399

v.) Accuracy = a+d/ (a+b+c+d) = 11/14 = 0.7857

b.) Techniques to handle class imbalance problems:

i.) **Alternative Metrics:** Here, we consider the rare class as more important (assigning it more weight) and use metrics like Precision, Recall and Weighted Accuracy to measure effectiveness of classification.

ii.) **Receiver Operating Characteristic Curve (ROC):** The TPR(true positive rate) and FPR(false positive rate) are plotted along the y-axis and x-axis respectively, a good classification model should lie somewhere on the upper left corner of the graph and the random classifier along the diagonal.

iii.) **Cost-Sensitive Learning:** Costs are calculated for various models using the cost matrix and the costs in turn can be used to make important decisions regarding the information.

iv.) **Sampling Based Approaches:** There are three techniques used undersampling (where lesser samples from majority class are taken), oversampling(where more samples of rare class are created) or hybrid sampling(which is a combination of both) for the training set, thus ensuring equal representation during the training phase.

3.)

a.) Maximum number of Association rules: (minsup>=0) $3^d - 2^{d+1} + 1 = 3^6 - 2^7 + 1 =$ **602**

b.) Maximum size of frequent itemsets: (minsup>0)

for all 6 itemsets individually ${}^6C_1$-**6**

now pairs ${}^6C_2$ – 15 possibilities – {Milk,Bread} , {Milk,Butter} ,{Milk,Cookies} ,{Milk,Beer} , {Milk, Diapers}, {Bread,Butter}, {Bread,Cookies}, {Bread,Diapers}, {Butter,Cookies}, {Butter, Diapers}, {Beer,Diapers} , {Beer,Cookies} ,{Diapers,Coookies} -**13**

For trios ${}^6C_3$ – 20 possibilities – {Milk,Beer,Diapers}, {Bread,Butter,Milk}, {Milk,Diapers,Cookies}, {Bread,Butter,Cookies}, {Beer,Cookies,Diapers}, {Milk,Diapers,Bread}, {Milk,Diapers,Butter}, {Diapers,Bread,Butter}, {Bread,Butter,Diapers} – **9**

For 4 ${}^6C_4$ – 12 possibilities – {Milk,Diapers,Bread,Butter} - **1**

For 5 and 6 we can see that there are 0 itemsets.

Total = **29**

c.) Maximum number of size-3 itemsets: (minsup>=0) ${}^6C_3$ = **20**

d.) Confidence of the rules {Bread -> Milk} and {Milk -> Bread}: c= sigma(X,Y)/sigma(X)

c{Bread->Milk} = 3/5 = **0.6**

c{Milk->Bread} = 3/5 = **0.6**

e.) Conditions under which rule {a->b}, {b->a} have the same confidence:

Since the numerator of confidence does not change for these two cases, it all depends on the denominator i.e. sigma(X), so basically:

-When ‘a’ and ‘b’ have same support count.