Name: Tharani Maarecivaanan
ID: 2575198

1. Pseudo-code for Genetic Algorithm

Genetic-algorithm is population-based algorithm inspired by natural evolution

(1) Initialize the population (randomly)

(2) Compute the fitness for each chromosome

(3) Pick the best-ranked individual to mate & replenish the population by applying

(a) Cross-over operator

(b) Mutation operator

(4) Go to step-2 and & repeat until the termination condition

2. Motivation behind the development of Random-forest algorithm.

→ The essential idea of bagging is to average many noisy but unbiased model & hence reduce its variances.

→ Trees are ideal candidate for bagging because they can capture the complex interactions in data. And when they grow deep, they have low bias.

→ Trees are also very noisy & they can averaged.

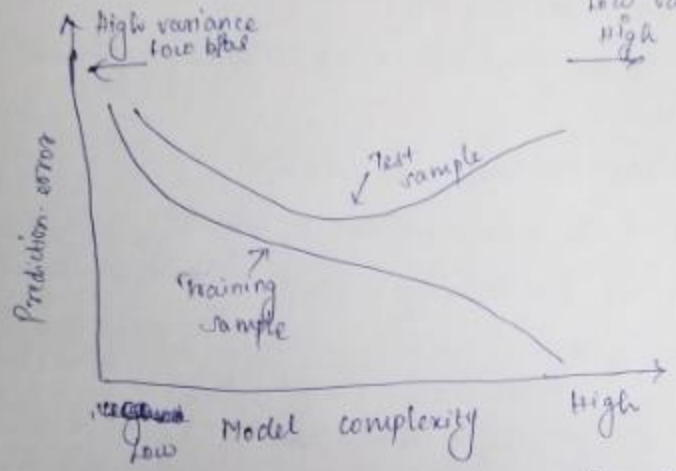
→ When the trees are generated in bagging they are ideally distributed.

→ So the expectation of average of B such trees is same as the expectation of average of each tree. Also the bias of bagged tree is same as the bias as ones & the only hope of improvement is ~~be~~ variance reduction.

→ Random forest algorithm is developed so that it can reduce the

variance without disturbing the bias.

3. (a) figure to show relationship among model complexity, prediction error, bias, variance



(b) The relationship btwn MSE (Mean square error) & RSS (Residual sum of squares) is given by,

$MSE = \frac{1}{N} RSS$ where N is the actual & predicted data pairs in RSS.

(c) Relationship btwn MSE, bias-error & variance error is given by,

$$MSE = (\text{bias-error})^2 + \text{variance-error}$$

4. When the utility matrix is given as

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	1			5	1		
B	5	5	4				
C			2	1	5		
D		3				3	

(2)

To compute the normalized rating, subtract each rating from average rating of that user.

So the utility matrix will become

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4/3			5/3	-7/3		
B	1/3	1/3	-2/3				
C			-5/3	1/3	4/3		

Now find the cosine distance between the vectors.

For eg: $sim(A,B) = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{4/3 \cdot 1/3}{\sqrt{(4/3)^2 + (5/3)^2 + (-7/3)^2}\ \sqrt{(1/3)^2 + (1/3)^2 + (-2/3)^2}}$

$$sim(A,C) = \frac{5/3 \cdot 1/3 + (-7/3) \cdot (4/3)}{\sqrt{(5/3)^2 + (4/3)^2 + (-7/3)^2}\ \sqrt{(-5/3)^2 + (1/3)^2 + (4/3)^2}}$$

for $sim(A,B)$ = it gives +ve value

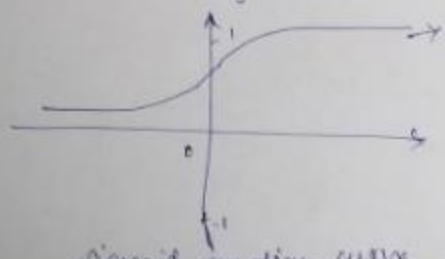
$sim(A,C)$ = it gives -ve value

It shows that eventhough A, C have watched 2 movies in common but they have difference in opinion.

And A & B watched one movie in common & rated similarly

Thus normalized cosine rating helps in finding the similarity measure.

5. characteristic curve (a) sigmoid function

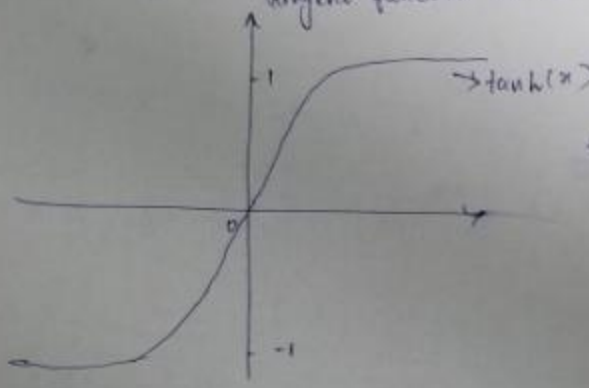


$$sin(x) = \frac{1}{1+e^{-x}}$$

sigmoid function curve

(b) hyperbolic tangent function

tangent function curve



$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

the start

6. Pros of content-based recommender sys

→ No need for data of other users so there will be no cold-start or sparsity problem.

→ Able to recommend new or unpopular items (No first rater problem).

→ Able to recommend to users with an unique taste.

→ Able to provide explanations by listing the features of an item & why it is recommended.

Cons

→ Not able to find appropriate feature (for images, videos).

→ Overspecilization

Unable to recommend an item outside the user's profile because user have many preferences.

Unable to exploit the quality judgement of other users.

→ Do not know how to build profile for new users & what to recommend for them.

7. Vector - equation for given ANN.

$$z^{(1)} = x$$

$$\eta^{(2)} = \beta^{(1)T} z^{(1)}$$

$$z^{(2)} = f_{sig} \cdot \eta^{(2)} \text{ & add } z_0^{(2)}$$

$$\eta^{(3)} = \beta^{(2)T} z^{(2)}$$

$$z^{(3)} = f_{sig} \cdot \eta^{(3)} \text{ & add } z_0^{(3)}$$

$$\eta^{(4)} = \beta^{(3)T} z^{(3)}$$

$$z^{(4)} = \hat{y} = f_{sig} \cdot \eta^{(4)}$$

8. Steps for k-fold cross validation for model selection.

(1) Swallowfully split the dataset D into k disjoint subsets $D_1, D_2 \ldots D_k$.

(2) For Vi, model M_i is

for j = 1 to k

(1) Train the model M_i using data $(D - D_j)$ & get the predictor P_{ij}

(2) Test P_{ij} & get the error E_{ij}

End for j

E_i = average error of E_{ij} which is the generalized error for model M_i.

(3) Find the best model M_i with the lowest generalized error E_i

(4) Retrain model M_i = best using full dataset D.

9. steps used for prediction using SVM

(1) convert data into the format of SVM package

(2) Conduct simple scaling on the data.

(3) Consider the RBF kernel $K(x, y) = \exp(-\gamma \|x - y\|^2)$

(4) Use the cross-fold validation to find parameters C & γ.

(5) Use the best parameters C & γ & train the dataset.

(6) Test.

10. Non-deterministic algorithm

(1) Artificial Bee colony

(2) Bat Algorithm

(3) Firefly Algorithm

(4) Monkey Algorithm

(5) Social spider Algorithm

(6) Cuckoo Search Algorithm