Name: Tharani
Maaneeivaannan
ID: 2575198

1.(a) Unsupervised learning

→ It is also known as 'learning without teacher'.

→ In unsupervised learning, the samples are given without their categories or outcomes.

→ The dataset X is given & the distribution is $Pr(X)$. The goal is to find the deviation from the distribution without any help & report if there is any error calculated. There is no measure of success.

(b) K-mean clustering algorithm

Input: dataset D $\{x_1, x_2, \ldots x_N\}$. parameters k. where $i = 1, 2, \ldots N$

(1) Initialize centroids $m_1, m_2, \ldots m_k$ randomly

(2) Compute the cluster assigment $c(i)$

$$c(i) = \arg\min_{1 \le k \le K} \| x_i - m_k \|^2$$

(3) compute

$$m_k = \frac{\sum_{i=1}^{N} I(c(i) = k)\, x_i}{\sum_{i=1}^{N} I(c(i) = k)}$$

where I is the indicator function. $I(c(i) = k) = 0$ for any k then change the m_k randomly.

(4) Go to step-2 & repeat until $c(i)$ do not change.

2(a) Longtail phenomenon

The distinction btwn physical world & the outside world is called as long tail phenomenon. The physical world has limited space & it cannot tailor the items based on ~~the~~ each customer. The long tail phenomenon forces the online world to recommend based on each customer preferences.

(b) Classification of reccomender s/m.

(i) content - based s/m.

It recommends based on the features/ properties of the item.

Eg:- If Netflix user watched a movie in "cowboy" genre, then it recommends a movie that is in the database from "cowboy" genre or style.

(ii) collaborative filtering s/m

It recommends ~~eyes~~ an item based on the item purchased by similar user.

It recommends an similar item purchased by ~~similar~~ users who rated items. ~~similarly~~ similarly.

3 (a) Bagging

~~Bagging is the tool for assessing statistical accuracy~~

Bagging is the bootstrap aggregation. Suppose we fit a model for a set of training data. It is denoted by $Z = \{(x_1, y_1), (x_2, y_2), \ldots \ldots (x_N, y_N)\}$. the predicted value will be $\hat{f}(x)$ at input x.

Bagging averages this prediction over a collection of bootstrap samples & hence reducing its variance.

for each bootstrap model Z^{*b}, we fit our model ~~[illegible]~~ & find the prediction $\hat{f}^{*b}(x)$.

Bagging estimate $\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{*b}(x)$

(b) Random forest for regression & classification

(1) Draw the bootstrap sample Z^* of size N from the training dataset.

(2) Grow the ~~trees~~ randomforest T_b for the bootstrap data by recursively repeating the following steps ~~by~~ for each terminal node of the tree until the minimum node size n_{min} is reached.

(a) select m variables randomly from the p variables.

(b) Pick the best variable / split-point among them.

(c) split the nodes into 2 daughter nodes.

(3) Output the ensemble of trees $\{T_b\}_1^B$

To make the new prediction at point x,

Regression: $\hat{f}_{rf}^{B} = \frac{1}{B} \sum_{b=1}^{B} \{T_b(x)\}$

classification: Let $\hat{C}_b(x)$ be the class prediction of the b^{th} random forest tree.

$\hat{C}_{rf}^{B}$ = majority vote of $\{\hat{C}_b(x)\}_1^B$

4. Pros ~~of~~ content_based recommender s/m.

→ No problem of cold-start & sparse data.

→ can recommend new ~~data~~ item

→ can provide detail because the recommendation is based on the ~~feature~~ features/properties of item.

Cons of the content-based system

→ do not know how to recommend the item that is unpopular.
~~to~~ the new ~~item~~ ~~recommendation~~

→ do not know how to recommend for new user.

→ overspecification

Let us consider the vectors given.

A	0	1	1	0	1	1	0	1	3α
B	1	1	0	1	0	1	1	0	4α

The dot product will be $2 + 4\alpha \cdot 3\alpha = 2 + 12\alpha^2$. The length of the vector will be $\sqrt{5+(3\alpha)^2} = \sqrt{5+9\alpha^2}$ and $\sqrt{5+(4\alpha)^2} = \sqrt{5+16\alpha^2}$

The cosine similarity between them will be,

$$\cos(A,B) = \frac{2+12\alpha^2}{\sqrt{5+9\alpha^2}\sqrt{5+16\alpha^2}}$$

Let us assume $\alpha = 1$, the movie rating as it, then cosine distance will be 0.81

Let the $\alpha = 2$, double the rating, then cosine value will be 0.94.

There is no much difference between them.

But when $\alpha = 1/2$, then the value will 0.6. It varies so much.

So the value depends on the scaling factor for numeric features of the vector.