**Date**: 17/08/2020

AI Assignment – 1 – 64 (TYCS)





**Soln :-**

Firstly, in Data Analytics, there exists two types of analysis that are

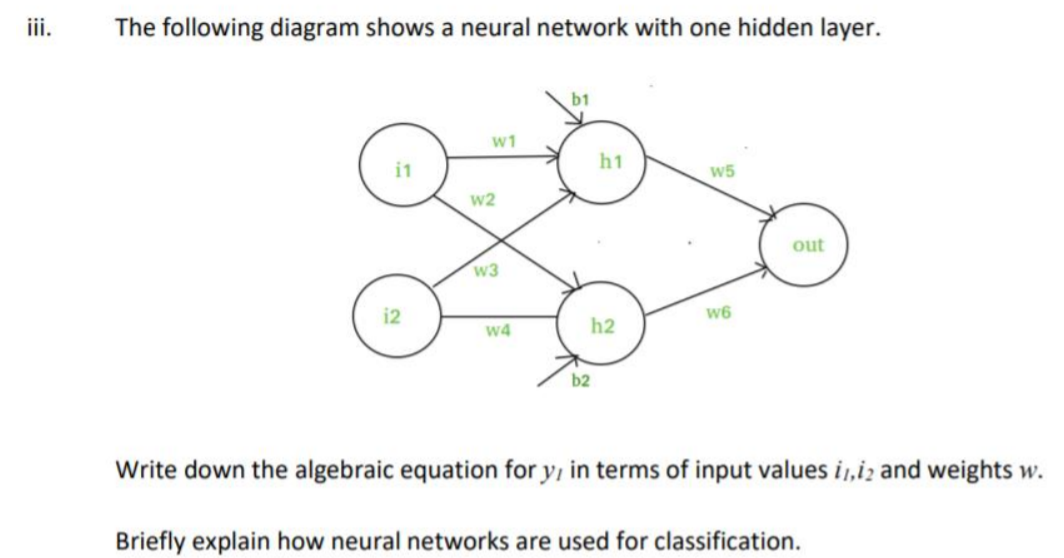
* **Classification** is the process of identifying the category or class label of the new observation to which it belongs.
* **This method is used for non-temporal or non-continuous data.**
* **Classification is used for differentiating and grouping the data into classes based on their behaviour/parameters.**
* **Classification can be binary classification or multi-class classification**
* **Example: Classification can be for Fraudulent/Non-Fraudulent customers.**
* **Prediction** is predicting a missing/unknown element of a dataset.
* Prediction plays an important role in analysing certain outcome from parameters that are continuous/ temporal in nature.
* Prediction cannot be used on Categorical Data.
* Prediction is just calculating meaningful information from a huge data set.
* **Prediction can be Uni-Variate or Multi-Variate prediction.**
* **Example: Prediction can be used for tabular data analysis.**



**Soln :-**

**Classifiers:**

* Logistic Regression: Cats Vs. Dogs
* Naïve Classifier
* Decision Trees
* Perceptron



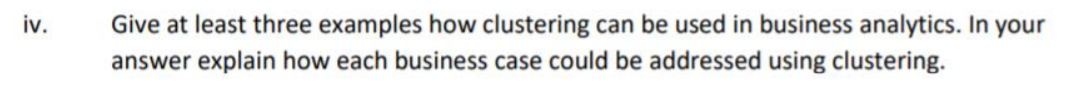
**Soln :-**

H1 = (W1 . I1 + W3 ­­. I2 + B1)

H2 = (W2 . I1 + W4. I2 + B2)

Y1 = (H1. W5 + H2.W6)

* Neural Networks are fundamentally linear considering the dot product of weights and inputs from previous layers. The added feature is the non-linear activation functions.
* This non linearity in neural networks permits the algorithm to model complex relationships beyond the scope of linear functions.
* The final layer – the output layer; in case of a classification model, returns a probability distribution signifying how much confidence the network has in classifying the data.
* Furthermore, loss functions, and Back-Propagation is used to train the neural network over a given dataset to get it better in the classification, by altering its weights by some proportion given by the loss function.
* The difference is then carried back from the output layer to the previous layers to modify the weights, thereby “training” the neural network to better its precision for the future



**Soln :-**

1. **Customer Segmentation:**

In the context of customer segmentation cluster analysis is the use of a mathematical model to discover groups of similar customers based on finding the smallest variations among customers within each group. The goal of cluster analysis in marketing is to [accurately segment customers](https://www.optimove.com/download/guide-advanced-customer-segmentation) in order to achieve more effective customer marketing

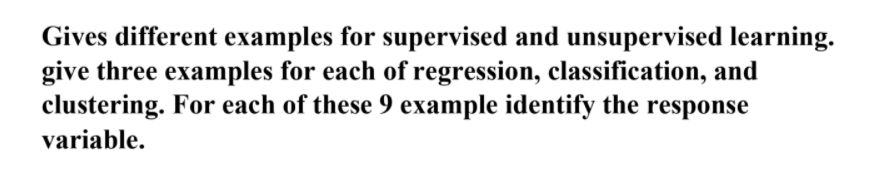
1. **News feed Generator:**

In context of Google Discover, that arranges and provides you a news feed of your preferences automatically. It uses a mathematical model having to discover similar topics of news and then makes clusters which are provided for preferences to users.

Accordingly, the news feeds are sent to that customer. Google’s name of this algorithm is called **Google Page Rank.**

1. **Marketing**:

Finding groups of customers with similar behaviour given a large database of customer data containing their properties and past buying records.

****

**Soln :-**

(R.V.: Response Variable)

**Regression:-**

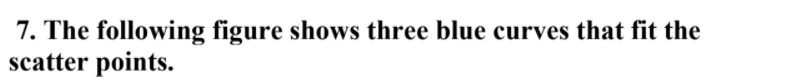
1. Salary Predictor R.V.: Experience
2. COVID Analysis R.V.: Daily Positivity Rate
3. Rainfall Predictor R.V.: Temperature

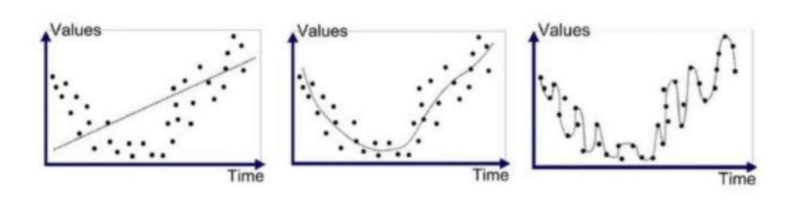
**Classification:-**

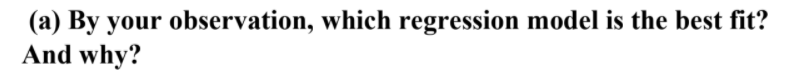
1. Email Spammer R.V.: Subject of Mail
2. Cat Vs Dog R.V.: Pixels
3. Digit recognizer R.V.: Pixels

**Clustering:-**

1. Customer Segmentation
2. News Feed Generator
3. Can be used for deciding new location of a restaurant.

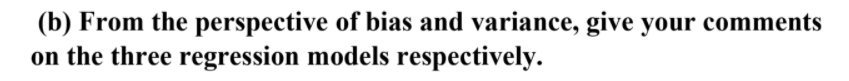
****

****

****

**Soln :-**

In the given figures, we can consider Model 2 as Best Fit because, the model function has too much complexity (parameters) to fit the true function correctly.



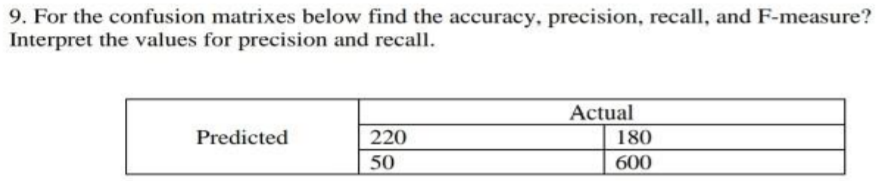
**Soln :-**

|  |  |  |
| --- | --- | --- |
| **MODELS** | **BIAS** | **VARIANCE** |
| **MODEL 1** | Very high Bias | Low Variance |
| **MODEL 2** | Optimal Bias | Optimal Variance |
| **MODEL 3** | Zero Bias | High Variance |



**Soln :-**

**Overfitting** happens when our model captures the noise along with the underlying pattern in data. The **problem** is that these concepts do not apply to new data and negatively impact the models ability to generalize.

****

**Soln :-**

Considering the following Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | **Actual Positive** | **Actual Negative** |
| **Predicted Positive** | 220 | 180 |
| **Predicted Negative** | 50 | 600 |

TP = 220

FN = 180

FP = 50

TN = 600

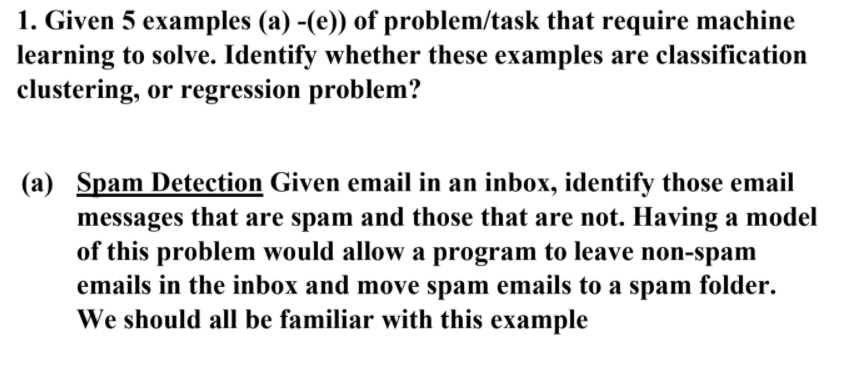
**Accuracy** = (TP + TN) / Total = (220 + 600) / 1050 = 0.78905

**Precision** = TP / ∑ Predicted Positive = 220 / 400 = 0.55

**Recall** = TP / ∑ Actual Positive = 220 / 270 = 0.8184

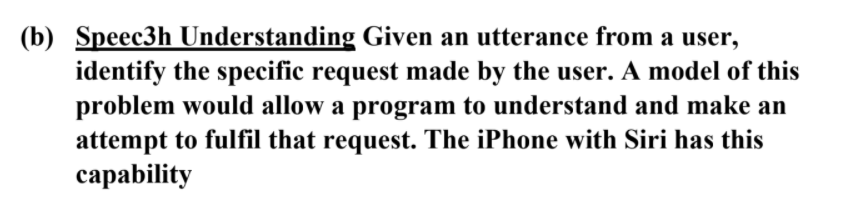
**F – measure** = (2 \* Precision \* Recall) / (Precision + Recall)

= (2\*0.55\*0.8148)/(0.55+0.8148) = 0.6567

****

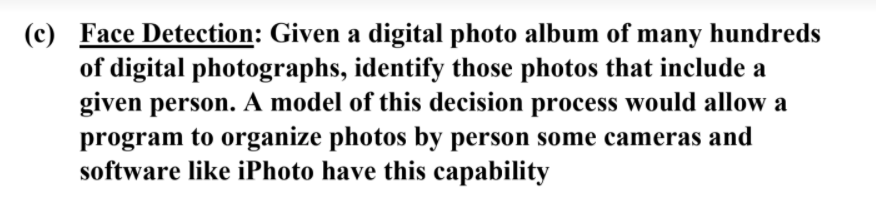
**Soln :-**

Classification.



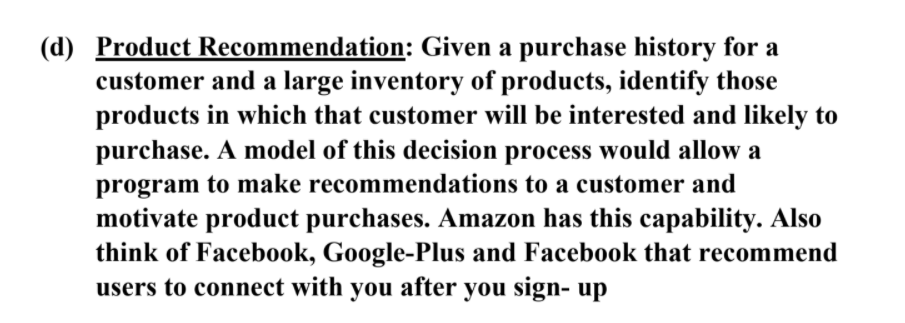
**Soln :-**

Classification. Sometimes Clustering can also be used for efficient replies.



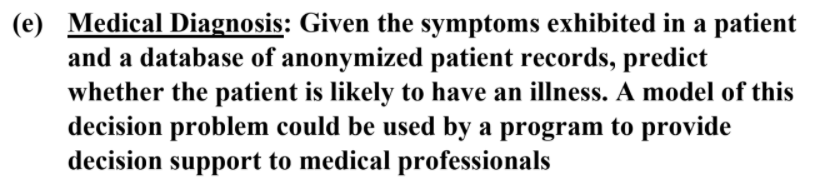
**Soln :-**

Classification



**Soln :-**

Clustering Algorithm



**Soln :-**

Classification and if data is parameterized then Regression can be used.