1.

a)

i. Facial Recognition

Recognize if a face matches a labelled set of facial features/faces or not, it is pattern recognition yes or no question and hence a **classification** problem

ii. Mobile Price Prediction

Predict the price of a unknown mobile device as a concrete real number and hence a **regression** problem

iii. Credit Card Fraud Prediction

Output belongs to ‘fraud’ or ‘not fraud’, again a true or false question and hence a binary **classification** problem

iv. Customer Churn Prediction

Output will be if a customer will churn (leave) or not, yes or no, binary **classification** problem

b)

A hypothesis producing a real, continuous and finite valued number is regression. The prediction in regression will be a tangible, concrete mathematical real number. The output variable/prediction is dependent on the input feature vector.

A hypothesis differentiating the input into labelled classes is classification. The prediction in classification will be whether the value belongs to a particular labelled class.

A hypothesis separating the input into different related groups that have no label is clustering. In clustering, we do not know what related clusters or groups of data are present in the input and neither do we know what the prediction result is going to be classified into.

Regression and Classification are supervised learning and clustering is unsupervised learning.

Supervised learning is when we teach the computer what to do and unsupervised learning is when we let the computer learn what it needs to do.

Examples:

Regression – Having trained a model to predict the gold price given a historical set of data where price is dependent on various attributes, predict today’s gold price.

Classification – Having trained a model to distinguish between various geometrical shapes, predict and classify as to what shape a new item belongs (circles, triangles, etc.).

Clustering – Given a data set for various genomes/DNA samples, group the data into clusters of similar values. Having done that, when a new genome sample is provided, match the sample to the best fitting cluster and predict to which category/cluster the new sample belongs.

2.

2019AIML618.ipynb

3.

2019AIML618.ipynb

Interpretation to business:

The salary offered to employees varies linearly with the years of experience of the employee. The salary for fresh graduates starts around 26,000 and goes up to 1,20,000 for someone with 10 years work-ex.

With the model we have come up as depicted, we can almost accurately predict the salary to be offered to employees/potential candidates when we know their years of experience.

The confidence of such a prediction is 95.7%.