|  |  |
| --- | --- |
| Index | Summary |
| 1 | * Machine learning is the most hit product of AI * In ML, algorithms are mostly based on Artificial Neural Networks * ANNs are called Deep Learning Based Framework |
| 1.1 | There are three types of ML   * Supervised (data with labels) and Unsupervised (data with no labels) * Online (collect and train on the spot) / batch (collect all data and then train, system will NOT update itself when it gets new data) * Instance-based (Line by line learning individually e.g. Row in Excel Sheet) / model-based (Train using all available data and the model will predict next value) |
| 1.2 **3:10** | In Supervised learning Labels are available with data. There are two major applications of Supervised Learning   * Classification: put a Label of some class that input data belongs to. There are two types of classification   + Binary   + Multiclass (Multi-label acc to Sir?) * Regression: (Calculate a number) put a Numerical Value on the data based on some parameters in data |
| 1.3 | In Unsupervised Learning, no labels are given with data and system learns automatically. Most common types of learning are   * Clustering (System automatically put similar instances in a group) * Visualization (dot plots: check relationship among data based on distance in x-y plane) and Dimensionality Reduction (System automatically decides what data points or features are important) * Associate rule learning (Market Basket Analysis: Close relationship of data or features?) |
| 1.4 | In Semi-supervised Learning, lots of data is unlabeled and some data is labeled. There are two ways of applying Semi-supervised Learning.   * Most Commonly, we run the classifier on unlabeled data for clustering, Association and so on. Then run on labeled data. We will not get result instantly * We first train the classifier on labeled data and then apply it on unlabeled data. One benefit is we will get results instantly even in the first run |
| 1.5 | Reinforcement Learning is the learning technique to maximize the reward and minimize penalty of a software agent. In machine learning, we increment the reward score of classifier when results are good and increment penalty when results are not good.  We also need to observe if data is relevant to output of classifier i.e. if there exists a relationship between data and output labels |
| 1.6 | In Batch or Offline Learning, we provide data with labels all in once to train the classifier.  It is computationally less expensive but when we get new data, we have to train the classifier on all the data available again to get better results. In Online Learning, System learns from data on the spot incrementally. It gives better results but computationally more expensive than Batch Learning |
| 1.7 |  |
| 1.8 |  |
| 1.9 | Model Selection is the most important step in MLDLC  There are several steps in MLDLC such as   1. Data acquisition    * Requirement Analysis    * Dataset Collection    * Dataset Analysis 2. Model Selection    * Model choice (based on type of data) (supervised / unsupervised etc)    * Model training    * Performance measurement    * Test Deployment 3. Error Analysis    * Error Analysis techniques    * Error Reduction    * Final Deployment   A model is also known as a classifier |
| 2 | There are set of **Rules** in every programming language that come from **Human Expert Knowledge** known as **Conventional Programming**.In machine learning, we don’t provide Human Expert Knowledge explicitly. |
| 3 | Whenever we go to solve a real world problem using **Traditional Approach**, we have   * Input   + Study the problem * Processing   + Write rules   + Evaluate   + Analyze Errors * Output   + Launch if no errors   In **Machine Learning** approach, we train the ML algorithm to write rules. To do that, we need large high quality data. Both data and rules can be updated dynamically in contrast to traditional approach  A machine learning algorithm may have two goals   * To solve the problem while providing high accuracy * To make humans better understand the problem |
| 4 |  |
| 5 | Deployment phase in MLDLC is same as SDLC   * Validation data is subset if test set data * Validation Set: A set of examples used to tune the parameters of a classifier, for example to choose the number of hidden units in a neural network * Test Set: A set of examples used only to assess the performance of a fully-specified classifier |
| 6  3:40 | Overfit: Classifer perform well on training data but not on testing data.  Underfit: Classifer doesn’t perform well on training data and testing data as well.  Confusion matrix is also applicable on multi-label data.  To reduce error   1. Change the parameters of the model 2. If error still persists then increase the amount of training data 3. If error still persists then change the ML model 4. If error still persists then Investigate the data and remove some irrelevant instances or attributes of instances   Regularization: Use samples of data instead of using whole data |
| 7 | **Training And Performance Measurement in Machine Learning**  Lets say we have a data of 100 faces with labels   * Separate out 90 of them and call this set as **Training data** * 10 of them called **Validation data** * Data from real world i.e. data without labels is known as **Test data** |
| 8 | **Main Challenges of Machine Learning**  Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entity mentioned in unstructured text into pre-defined categories such as person names, organizations, locations |
| 9 | **End to End kering in Machine learning**  There are several libraries used in machine learning such as psychic learn, numpy and tensorflow etc. |
| 10 | **SRL Frams the problem** |
| 11 | **Data Pipeline** |
| 12 | **Data acquisition** |
| 13 | **Data preprocessing** |
| 14 | **Training & Testing Data** |