|  |  |
| --- | --- |
| Index | Summary |
| 1 | **Introduction to Deep Learning**  \*ML or DL has one common purpose that is *to introduce perception or mimic the perception abilities of humans.*  \*Perception: the ability to become aware of something through the senses.  \*Vision: How human understands what he sees.  \*Speech: Communication using vocalization, deliver message by speaking.  \*Ear (Speech Recognition): Receive and understand the spoken message.  \*Text: Understanding the written message.  \*Deep Learning have improved the above-mentioned perceptions for a computing device. |
| 2 | **Importance of Data for Deep Learning Algorithms**  \*In modern world, data is considered as oil.  \*Algorithms of AI such as face identification, recognition, smile identification, action identification and so on, is done well with the help of data.  \*If we have 5000 instance or rows of data of a particular category then we can prepare a classifier using this data.  \*When instances of labeled data such as images approaches to 10million then computer gain quality extent of perception approaching the human perception.  Annotated data: Labeled data |
| 3 | **Interoperability of Deep Learning Algorithm**  \*All the big commercial enterprise such as google, Facebook etc. are investing in Deep Learning because these companies are oriented about the behavior of user. They apply AI algorithms on the user data to make user experience better. E.g. Facebook 10yr challenge, recommendations, friend suggestion etc.  \*Deep learning performs well on small as well as complex problems. All we need is lots of data.  \*Interoperability: Same DL algorithms, tools or techniques are applicable on various problem domains with a minor change algorithm. E.g. the same algorithm can be useful for speech recognition, Bioinformatics, audio-processing or weather data. |
| 4 | **Deep Learning Frameworks**  In deep learning framework, there are 4 blocks   1. Classification (to give classification label to data)    1. Choose a classifier such as Logistic Classifier    2. Stochastic optimization: reduce errors and improve accuracy    3. Data and parameter tuning: Fine tune the data features 2. Deep Networks (multiple layers performing individual task such as eye detection, lips detection is combined to make a face identifier)    1. DNs: Collaborate layers that perform task of similar level    2. Regularization: Understand how sub-problems will solve the parent problem 3. Computer Vision    1. Convolutions    2. Neural Networks 4. Text Processing (or any time series data)    1. Recurrent Neural Networks or RNNs    2. Embeddings    3. Longest Short-term memory or LSTM |
| 5 | **Linear Classification**  \*Classification: Assignment of specified or categorized labels to data E.g. digit recognition.  \*Logistic classifier is an S shaped binary function that classifies the given data between 0 and 1.  \*W stands for **weights.** B stands for **Bias** and it is our prior knowledge**.** X is **input vector** or **input class.** Y is **output vector.**  \*Wx + b yield scores as output vector.  \*W and B come from training data.  \*Softmax function converts scores of output vector into proper probabilities i.e. sum of all entities in output vector is 1. |
| 6  2:40 | **One Hot Encoding**  \*After applying softmax function to assign proper probabilities to output vector entities, the maximum probability is set to ‘1’ and rest to ‘0’. It is called one hot encoding.  \*It is binary classification.  \* Scaling is the technique of standardizing the features of data within a defined range. It tells us that how a particular feature is comparatively significant as compared to other features. |
| 7 | **Entropy** |
|  |  |
| 12  2:20 | **Linear Model and Its Advantages**  There are two models used in machine learning   * Linear Model:   + Differentiate between classes using straight line.   + Can’t handle real life problems.   + Let’s break down WX + B = Y into discrete blocks.   + X is input vector that contains N inputs.   + Y is output vector that contain K outputs.   + Total number of possible parameters is (N+1)K where 1 signifies the corresponding missing value. Benefits of Linear model are:     - They have Controlled number of parameters. Due to this limit, linear model is not quite useful in real life problems. Also due to this limit, linear model is simpler and easier to handle.     - They are additive. Linear model should qualify the straight line. It will fail if we multiply two variables e.g. x\*x parabola not a linear model.     - They are stable i.e. they show stability on even low value of X so Y. Linear models are comparatively more stable. i.e there is one-to-one correspondence among X and Y.     - Their derivates are stable and constant i.e. are defined in fixed range.     - They are less complex. So, computation is easy. * Non-Linear Model |
| 13 | **Deep Learning Architecture**  \*To make a non-linear function, it should be the combination of linear functions.  \*we use linear functions because it is less computationally expensive, less complex, easy to handle.  \*Activation function is applied on linear data in order to combine linear functions.  \*The activation function in this module is RELU (Rectified Linear Unit). |
| 14 | **Deep Learning Algorithm | Chain Rule** |
| 15 | **Forward and Backward Propagation**  \*There are two types of propagation in ML   * Forward Propagation (Directed from left to right i.e. from X towards Y or Assignment of labels to input data) * Backward Propagation (Directed from right to left i.e. from Y towards X or To update weights in order to make ANN perform better. It is derivative based)   \*Propagation is stack of simple mathematical operations.  \* Backward propagation is more complex as it involves derivatives. It is 2 time more complex than forward propagation. |
| 16 | **Over Fitting and Under Fitting**  \*Overfitting: Model works on known data and can’t work on incoming unknown data.  \*Underfitting: Model can’t perform well on even known data. We need more data to overcome this problem. |