Unit 1 : Foundations of Deep Learning

Machine Learning

1. Subfield of AI developing theories of learning and building machines.
2. Learning → Gaining new symbolic knowledge and development of cognitive skills through practice.
3. Machine Learning definition : " A computer program is said to learn from experience E wrt some tasks T and performance measure P if its performance at tasks T as measured by P improves with E".
4. Data mining → Application of machine learning methods to large databases.
5. Goal of ML → Devise learning algorithms that do the learning automatically without human intervention.
6. ML is important due to:
    a. Relationships and correlations are hidden within large amounts of data so ML may be able to find these relationships.
    b. New knowledge about tasks are constantly discovered by humans.
    c. Helps find solutions to problems in CV, robotics, as it uses statistics to build models to find inferences.
7. Applications :
    a. NLP → Teach computers to understand, interpret and generate human language.
    b. Healthcare → Prediction of patient outcomes, identify potential outbreaks of diseases.
    c. Agriculture → Optimise crop yields and resource efficiency.

d. Self driving cars → Intelligent decisions about navigation and avoiding accidents.

e. Process of ML

   i. Data Collection → Collect data for training

   ii. Data preparation → Pre processing, normalisation, feature selection etc.

   iii. Model Selection

   iv. Model training

   v. Model evaluation

   vi. Model tuning

   vii. Deployment

Deep Learning

1. Deep refers to the number of hidden layers in the neural network.
2. Deep Learning → Subset of ML which is predicated on the idea of learning from example.
3. Deep learning → Repeated composition of functions can often reduce the requirements on the number of base functions by a factor that is exponentially related to the number of layers in the network.
4. In DL, computer model learns to perform classification tasks directly from images, text or sound. Models are trained by large set of labeled data and neural network architectures that contain many layers.
5. More layers enable precise results.
6. Consists of the following:

   a. Boltzman machines for unsupervised learning, GANs.

   b. Supervised learning, Convolutional neural networks → Image processing

   c. Recurrent neural networks → Allowing to train on processes in time

   d. Recursive neural networks → allowing to include feedback

Supervised and Unsupervised learning

1. Supervised → Input data is labelled, Unsupervised learning → Data is not labelled
2. Supervised learning → Model is trained on labelled data, meaning the data is provided with the correct output for each example in the training set.
3. Goal → Make predictions based on this input output mapping
4. Working of supervised learning
    a. Data prep
    b. Model selection
    c. Training
    d. Evaluation
    e. Tuning
    f. Deployment
5. Advantages:
    a. Wide range of tasks → regression, classification etc.
    b. Easily evaluated using performance metrics
    c. More accurate than unsupervised learning
    d. Well studied approach
6. Disadvantages
    a. Requires labelled data, can be time consuming
    b. Highly dependent on quality of data
    c. Prone to overfitting
    d. Sensitive to outliers
7. Unsupervised learning → Algorithm learns to identify patterns and structure in unlabelled data.

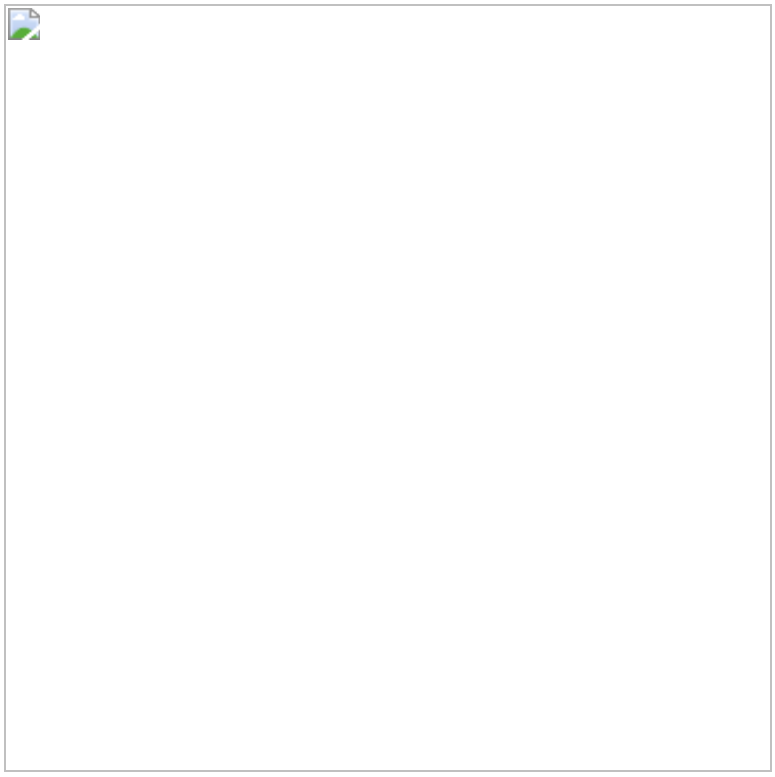
8. Advantages:
    a. Can identify patterns which arent immediately apparent to human
    b. Doesnt require labelled data
    c. Can cluster data into groups and identify anomalies.
9. Disadvantages:
    a. Difficult to predict as theres no target variable
    b. Sensitive to the quality of data
    c. Difficult to tune
    d. Computationally expensive
    e. Can be prone to identify irrelevant patterns in data

Bias Variance trade off

1. Bias → Discrepancy between our actual values and the predictions.
    - Bias is the error introduced by approximating a real-world problem with a simplified model.
    - A high bias model tends to oversimplify the underlying patterns in the data and may fail to capture the true relationships between features and target variables.
    - Models with high bias are often too simple and underfit the data.
    - IN SHORT → HIGH BIAS CAUSES UNDERFITTING
2. Variance → Variability of the models forecast for a certain data point or values, which indicates how widely distributed our data is.
    - Variance refers to the model's sensitivity to fluctuations in the training data.
    - A high variance model captures noise in the training data as if it were real signal, leading to poor generalization to unseen data.

- Models with high variance are overly complex and tend to overfit the training data.
- IN SHORT → `HIGH VARIANCE CAUSES OVERFITTING`

3. Aim → Low bias and low variance.
4. The bias-variance tradeoff suggests that as you decrease bias in a model (make it more complex), you increase its variance, and vice versa. The goal is to find the right balance between bias and variance to achieve optimal model performance.



Hyperparameters

1. They are variables whose values influence the learning process and define the parameter values.
2. Examples of hyperparameters:
    - Test to train ratio
    - Learning rate: Controls the step size during optimization.
    - Number of hidden layers: Determines the depth of a neural network.
    - Number of neurons in each layer: Affects the capacity of the model to learn complex patterns.
    - Activation functions: Determines the non-linear transformation applied to the output of each neuron.
    - Batch size: Specifies the number of samples processed before updating the model's parameters.
    - Regularization parameters: Control the degree of regularization applied to prevent overfitting.
    - Dropout rate: Specifies the fraction of neurons randomly dropped out during training to prevent overfitting.

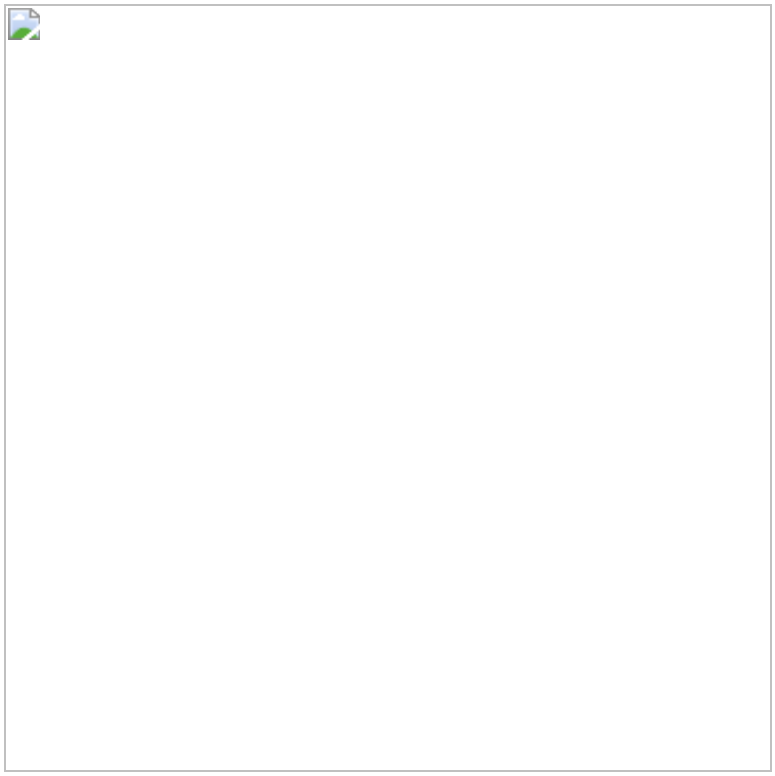
Underfitting / Overfitting regularisation

Underfitting

1. Underfitting → Model isnt able to recognise the underlying pattern in the data
2. It performs well on the training data but bad on the testing data
3. It happens when we try to develop a linear model but the data contains fewer non linear data.

Overfitting

1. When a model fails to produce reliable predictions on test data, its overfitted.
2. The model learns from noise and errorneous data when trained, thus yielding high variance. Due to too many details and noise, it fails to identify the data



Regularisation

1. Overfitting can be resolved if a method to lessen the complexity is found.
2. Complex models are penalised by regularisation.
3. When regularisation terms are included, the model seeks to reduce both loss and complexity of the model, and hence it decreases the variance without significantly increasing the bias.

Limitations of ML

1. Overfitting → refer above for desc
2. Lack of explainability → Many models are considered “black boxes” because its hard to understand how a model arrived at a particular decision.
3. Data bias → Models can amplify bias present in the data

4. Lack of diversity → If the training data is not diverse the model doesnt perform well on a diverse set of inputs
5. Requires large amount of data
6. Difficult to handle complex relationships in data
7. Security and privacy risks → deepfakes 🙂

History of ML

1. 1940s-50s → Researchers began developing models of neural networks
2. 60-70s → AI got alot of interest and funding, and experimentation of complex architectures
3. 80-90s → Foundations like backpropogation, multilayer neural networks and CNNs
4. early 2000s → Available power computers and data, more complex neural nets
5. 2010s → state of art perfomance, deep learning usage in academia and research

Advantages and Challenges of Deep Learning

Advantages:

1. Feature Engineering → Examines the data in search of complex features that combine them to enable faster learning
2. Best possible results for unstructured data → Different data formats can be used to train DL models.
3. No need for well labelled data → Algorithms are excellent at learning without any rules
4. High quality results → Can execute thousands of activities in a short time.

Challenges

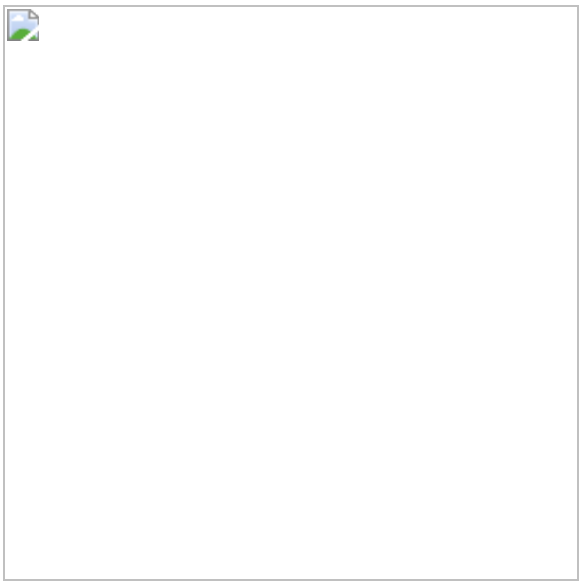
1. Deep neural networks require alot of training data
2. Includes bias related problems

3. Depending on the quantity and size of DL models, optimising computing expenditures is required.
4. Data security measures to ensure privacy.

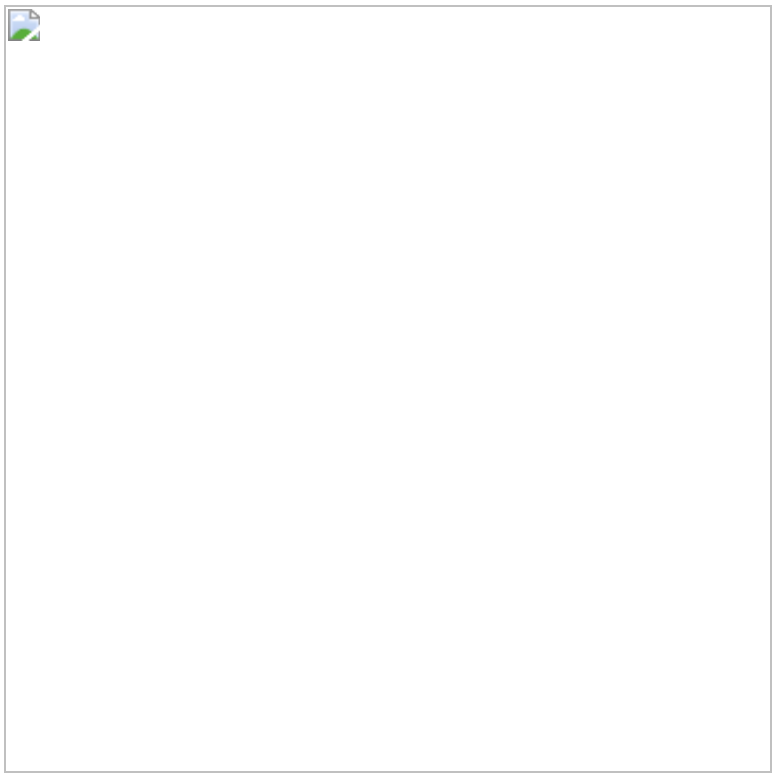
Understanding how DL works in 3 figures

(DIAGRAMS ARE IMPORTANT FOR THIS)

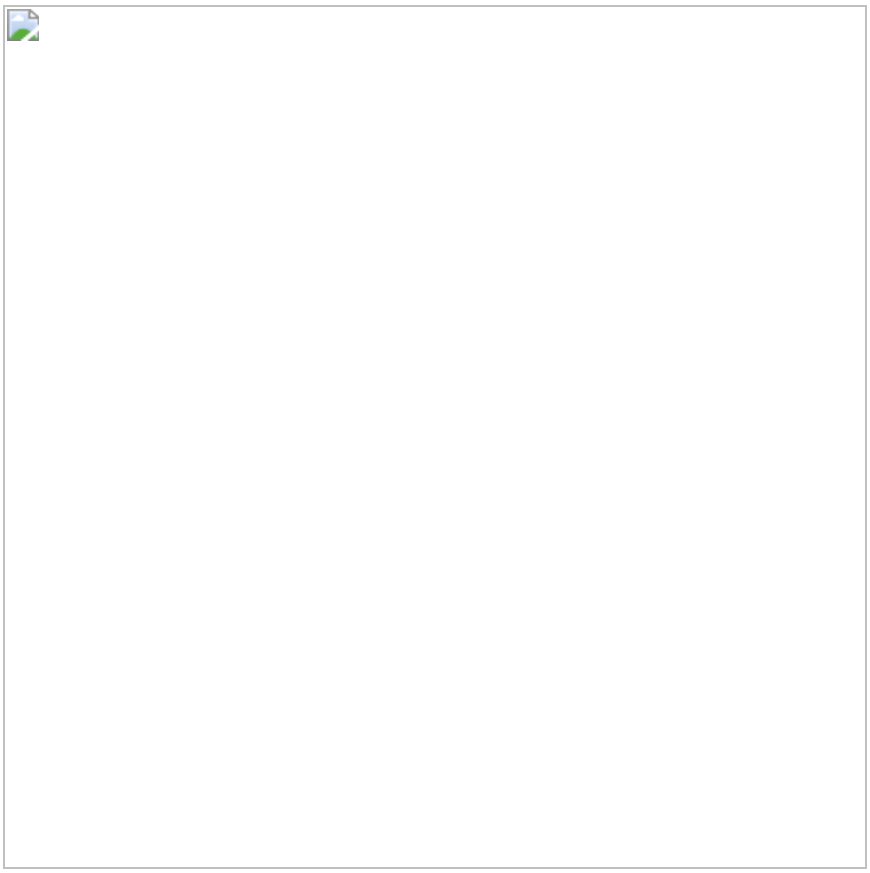
1. Neural network is composed of layers of interconnected “neurons”, inspired by the structure and the function of neurons in the brain.
2. Input layer recieves the raw data and output layer produces the final output of the network. Layers in between are called hidden layers



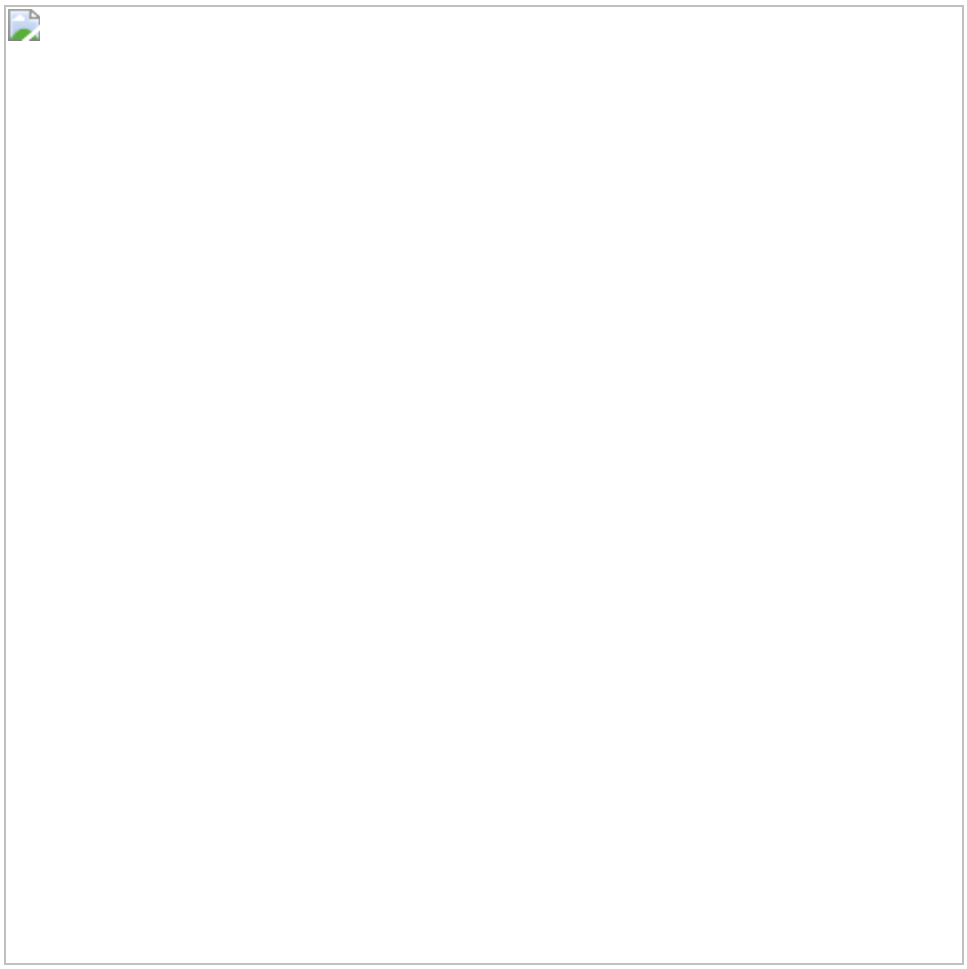
3. Training a neural network → The dataset is used for training the network, the network is presented with inputs and desired outputs.
4. This process is repeated for many examples in the dataset and it gradually learns to make accurate predictions on new, and unseen examples.



5. The last figure help understanding forward and backward propagation. Its the process of passing input data through the layers of the network, computing the output, and adjusting the weights in the backward pass by using optimisation algorithms like Gradient descent,



Common Architectural Principles of Deep Network



1. To create a model that classifies data highly accurately, we need more hidden layers to produce extremely complicated models.
2. A feed forward technique is used to generate output, after passing through the entire neural network.
3. Multiple layers → Increases the abstract representations of hidden data.
4. Non linear activation functions → ReLU or Sigmoid are used to introduce non linearity in the computation.
5. Gradient based learning → Adjusts the weights of the network to minimise the error on the training data.
6. Backpropogation → Allows the gradients of the error wrt to the weights to the network to be efficiently computed.

7. Dropout → Prevents overfitting, consists of randomly dropping out some nodes to effectively average over multiple different nodes.
8. Batch normalisation → Improves the stability and speed of training

Architecture design

Decisions to be made for designing the architecture of a NN:

1. Number of layers → This depends on the complexity of the task and the size of the dataset.
    a. Deeper network can learn more complex details but should have more data to prevent overfitting
2. Number of neurons in a layer → Capacity of the network.
    a. Larger neurons means more parameters to learn from the data, but also increases the risk of overfitting.
3. Type of activation function → They introduce non linearity into the network. eg. Sigmoid, TanH, ReLU, and leakyReLU.
4. Type of layer → Convolutional layers and recurrent layers are suitable for different types of data.
    a. Convolutional layers → Image data
    b. Recurrent layers → Natural language or time series
5. Regularisation strategy → Common techiques for regularisation are Dropout, weight decay and early stopping.

Applications of DL

1. Computer vision
2. NLP
3. Autonomous systems

4. Healthcare
5. Finance
6. Gaming
7. Agriculture
8. Self driving cars
9. Robotics and reinforcement learning

(elaborate karayche points example deun)

Introduction and use of popular tools for ML and DL

1. Tensorflow
    a. Open source
    b. Set of tools for building and deploying ML models
    c. Easy deployment for variety of models like mobile devices and browsers.
    d. Image and speech recognition, NLP and time series forecasting
2. Keras
    a. High level neural networks API
    b. Can evaluate DL models
    c. Compatible with Tensorflow and Theano
    d. Can be used to quickly build and experiment with DL architectures
3. Pytorch
    a. Open source ML library based on the torch library.
    b. Used for NLP and CV.
    c. Has dynamic computation graphs, for flexibility and debugging.
    d. Has strong visualisation and model interpretability tools.

4. Caffe
    a. Developed by Berkeley vision and learning center.
    b. Used for image classification and segmentation.
    c. Used for speed and efficiency, and for training large scale models like ImageNet.
5. Shogun
    a. ML library that provides a range of algorithms like classification, regression, clustering, dimensionality reduction.
    b. Implemented in C++ and provides interfaces to languages like Python and Matlab.