AI ML DS DL

AI: Programs that can sense, reason act and adapt.

General AI - Planning, decision making, identi fying objects, recogn izing sounds, social &
business transa ctions

Applied AI - driver less/ Autonomous car or machine smartly trade stocks

ML: Instead of engineers "teaching" or progra mming computers to have what they need to carry out tasks, that perhaps computers could teach themselves - learn something without being explicitly programmed to do so. ML is a form of AI where based on more data, and they can change
actions and response, which will make more efficient, adaptable and scalable. e.g., navigation apps and recomm end ation engines.

DS: Data science has many tools, techni ques, and algorithms called from these fields, plus others -to handle big data The goal of data science, somewhat similar to machine learning, is to make accurate predic tions and to automate and perform transa ctions in real-time, such as purchasing internet traffic or automa tically generating content.

DL: It is a ML Technique. uses large neural networks.. Teaches computers to do what comes naturally to humans.

SL USL RL

supervised learning:
In a supervised learning model, the algorithm learns on a labeled dataset, to generate reasonable
predic tions for the response to new data. (Forec asting outcome of new data)

- Regression
- Classi fic ation

Unsupe rvised learning:
An unsupe rvised model, in contrast, provides unlabelled data that the algorithm tries to make sense of by extracting features, co-occ urrence and underlying patterns on its own. We use unsupe rvised learning for

- Clustering
- Anomaly detection
- Associ ation
- Autoen coders

Reinfo rcement Learning:
Reinfo rcement learning is less supervised and depends on the learning agent in determ ining the output solutions by arriving at different possible ways to achieve the best possible solution.

Archit ecture of ML:
Business unders tan ding:
Understand the give use case, and also, it's good to know more about the
domain for which the use cases are built.

Data Acquis ition and Unders - tan ding: Data gathering from different sources and unders - tanding the

SL USL RL (cont)

> data. Cleaning the data, handling the missing data if any, data wrangling, and EDA(Exploratory data analysis).

Modeling:

Feature Engineering - scaling the data, feature selection - not all features are important. We use the backward elimination method, correlation factors, PCA and domain knowledge to select the features.

Model Training based on trial and error method or by experience, we select the algorithm and train with the selected features.

Model evaluation Accuracy of the model , confusion matrix and cross-validation. If accuracy is not high, to achieve higher accuracy, we tune the model...either by changing the algorithm used or by feature selection or by gathering more data, etc.

Deployment - Once the model has good accuracy, we deploy the model Once we deploy, we monitor the performance of the model.if its good...we go live with the model or reiterate the all process until our model performance is good. It's not done yet!!!

What if, after a few days, our model performs badly because of new data. In that case, we do all the
process again by collecting new data and redeploy the model.

Sponsored by CrosswordCheats.com
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Algos

Linear Regression:

Linear Regression tends to establish a relati onship between a dependent variab le(Y) and one or more indepe ndent variab le(X) by finding the best fit of the straight line.

The equation for the Linear model is Y = mX+c, where m is the slope and c is the intercept

OLS Stats Model (Ordinary Least Square):

OLS is a stats model, which will help us in identi fying the more signif icant features that can has an influence on the output. OLS model in python is executed as:

```
lm = smf.ol s(f ormula = 'Sales ~ am+con stant', data = data).f it() lm.con f_int() lm.sum mary()
```

What is Mean Square Error?

The mean squared error tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the "errors") and squaring them.

Why Support Vector Regres sion?

Difference between SVR and a simple regression model?

In simple linear regres sion, try to minimize the error rate. But in SVR, we try to fit the error within a certain threshold

Algos (cont)

> Logistic Regression:

The logistic regression technique involves the dependent variable, which can be represented in the binary (0 or 1, true or false, yes or no) values, which means that the outcome could only be in either one form of two. For example, it can be utilized when we need to find the probability of a successful or fail event.

Decision Tree:

A decision tree is a type of supervised learning algorithm that can be used in classification as well as regressor problems. The input to a decision tree can be both continuous as well as categorical. The decision tree works on an if-then statement. Decision tree tries to solve a problem by using tree representation (Node and Leaf).

Assumptions while creating a decision tree: 1) Initially all the training set is considered as a root 2) Feature values are preferred to be categorical, if continuous then they are discretized 3) Records are distributed recursively on the basis of attribute values 4) Which attributes are considered to be in root node or internal node is done by using a statistical approach.

How to handle a decision tree for numerical and categorical data?

Algos (cont)

> If the feature is categorical, the split is done with the elements belonging to a particular class.

If the feature is continuous, the split is done with the elements higher than a threshold.

Random Forest:

Random Forest is an ensemble machine learning algorithm that follows the bagging technique. The base estimators in the random forest are decision trees. Random forest randomly selects a set of features that are used to decide the best split at each node of the decision tree

Variance and Bias tradeoff:

Bias: It is the difference between the expected or average prediction of the model and the correct value which we are trying to predict. Imagine if we are trying to build more than one model by collecting different data sets, and later on, evaluating the prediction, we may end up by different prediction for all the models. So, bias is something which measures how far these model prediction from the correct prediction. It always leads to a high error in training and test data.

Variance: Variability of a model prediction for a given data point. We can build the model multiple times, so the variance is how much the predictions for a given point vary between different realizations of the model.

High Bias, Low Vrariance - Underfitting
High Variance, Low Bias - Overfitting

Sponsored by CrosswordCheats.com
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Algos (cont)

> A confusion matrix is a table that is often used to describe the performance of a classification model
(or “classifier”) on a set of test data for which the true values are known. It allows the visualization
of the performance of an algorithm.
True Positive Rate:
Sensitivity (SN) is calculated as the number of correct positive predictions divided by the total number of positives. It is also called Recall (REC) or true positive rate (TPR). The best sensitivity is 1.0, whereas the worst is 0.0.
True Negative Rate
Specificity (SP) is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called a true negative rate (TNR). The best specificity is 1.0, whereas the worst is 0.0.
KNN means K-Nearest Neighbour Algorithm. It can be used for both classification and regression.Also called an instance- based or memory-based learning

DL

What is perceptron and how it is related to human neurons?

DL (cont)

> If we focus on the structure of a biological neuron, it has dendrites, which are used to receive inputs. These inputs are summed in the cell body and using the Axon it is passed on to the next biological neuron.
kind of problem can be solved by using deep learning?
mage recognition, ☐ Object Detection, ☐ Natural Language processing- Translation, Sentence formations, text to speech, speech to text, ☐ understand the semantics of actions
Forward propagation: The inputs are provided with weights to the hidden layer. At each hidden layer, we calculate the output of the activation at each node and this further propagates to the next layer till the final output layer is reached. Since we start from the inputs to the final output layer, we move forward and it is called forward propagation
Backpropagation: We minimize the cost function by its understanding of how it changes with changing the weights and biases in a neural network. This change is obtained by calculating the gradient at each hidden layer (and using the chain rule). Since we start from the final cost function and go back each hidden layer, we move backward and thus it is called backward propagation
. Backpropagation is the fast, simple and easy to program.

DL (cont)

> ☐ It has no parameters to tune apart from the numbers of input.
☐ It is the flexible method as it does not require prior knowledge about the network
☐ It is the standard method that generally works well.
☐ It does not need any special mentions of the features of the function to be learned
Epoch – In the context of training a model, epoch is a term used to refer to one iteration where the model sees the whole training set to update its weights.
Regularization
❒ Dropout – Dropout is a technique used in neural networks to prevent overfitting the training
data by dropping out neurons with probability p > 0. It forces the model to avoid relying too
much on particular sets of features.
Remark: most deep learning frameworks parametrize dropout through the ’keep’ parameter 1−p.
❒ Weight regularization – In order to make sure that the weights are not too large and that
the model is not overfitting the training set, regularization techniques are usually performed on
the model weights. The main ones are summed up in the table below
Hyperparameter tuning in deep learning:

Sponsored by CrosswordCheats.com
Learn to solve cryptic crosswords!
http://crosswordcheats.com

DL (cont)

> The process of setting the hyper-parameters requires expertise and extensive trial and error. There are no simple and easy ways to set hyper-parameters — specifically, learning rate, batch size, momentum, and weight decay.
Approaches to searching for the best configuration:
• Grid Search
• Random Search
Transfer learning – Training a deep learning model requires a lot of data and more importantly a lot of time. It is often useful to take advantage of pre-trained weights on huge datasets
that took days/weeks to train, and leverage it towards our use case. Depending on how much
data we have at hand, here are the different ways to leverage this:
Learning rate – The learning rate, often noted α or sometimes η, indicates at which pace the
weights get updated. It can be fixed or adaptively changed. The current most popular method
is called Adam, which is a method that adapts the learning rate.
Adaptive learning rates – Letting the learning rate vary when training a model can reduce
the training time and improve the numerical optimal solution. While Adam optimizer is the
most commonly used technique, others can also be useful. They are summed up in the table

DL (cont)

> below:

QS

Why can we not use Multi layer perceptron for text data?
A multi layer perceptron assumes that the input is of constant size. In natural language, the length of the sentences can vary. There multi layer perceptron is not applic able.
Can we use a multi layer perceptron for regression purpose, how?
A multi layer perceptron can be used for a regression problem, it can be done by removing the activation function at the output node and using a suitable cost function.
What is the difference between Stochastic Gradient Descent and Batch
Gradient Descent?
Sto chastic Gradient Descent uses only one instance to compute the loss and update the parame ters. As a result, it converges faster but often yields a sub-op timal solution.
Batch gradient Descent on the other hand uses the whole data to compute the loss and update the parame ters. As a result, it converges at the slowest rate but guarantees an almost optimal solution.
What is the signif icance of 1 X 1 convol utions?

QS (cont)

> 1x1 convolutions are primarily used as a dimensionality reduction technique, it is primarily used to vary the number of filters in the convolution layers, it can be used to either increase or decrease the number of filters.
What is object detection and object localisation ?
Object detection is a process in which the model predicts whether the object is present
in the image or not.
In object localisation, the model outputs the coordinate values where the object is present within the image, given the object was present in image.
Explain the concept of dead unit?
A dead unit in a deep neural network is a neuron which is experiencing the vanishing gradient problem and covariate shift. In this state, the neuron learns extremely slowly or apparently stops learning altogether.
What do you mean by Learning rate?
Learning Rate is a factor by which the weights of neural network are updated in each cycle.
what is covariate shift and how is the problem solved?
Covariate shift is a condition when the neurons in deep neural networks stop learning or
learn extremely slow due to the vanishing gradient. It can be solved by (or combination of) following
Batch normalisation

Sponsored by CrosswordCheats.com
Learn to solve cryptic crosswords!
http://crosswordcheats.com

QS (cont)

> Careful Initialisation
Slow learning rate
dropout

In training a neural network, you notice that the loss does not decrease in the few starting epochs. What could be the possible reasons?

There are 4 possible scenarios where this is possible,

Learning rate is very low
Regularisation parameter is very high
Optimisation is stuck in the local optima
Optimisation started on a plateau

Dropout and DropConnect are both regularization techniques for Neural Network. Is there a difference between these two? How is setting dropout =0.3 different from drop connect =0.3?

The function drop-out in a layer assigns a probability p to every node in that layer such that, that node will not be included in the computation during the runtime with respect to probability p. (0.3 in question)

The function drop-connection in a layer is a probability p for every node in that layer such that, there is a chance p such that , that node will skip a connection to consecutive layer by the probability p.

What are the factors to select the depth of neural network?

A. Type of neural network (eg. MLP, CNN etc)

QS (cont)

> B. Input data
C. Computation power, i.e. Hardware capabilities and software capabilities
D. Learning Rate
E. The output function to map

What the problems with deep networks?

In case of very deep neural networks, the weights of the hidden layer might experience
either vanishing and explosive gradient problem. They are also prone to “overfitting” the data.
It is difficult to train a neural network and they take a very long time to train.

How are weights initialized in a neural network?

Usually weights can be randomly initialised to a random small value, but that can lead to
vanishing gradients and exploding gradient in case of Deep Neural Network. Therefore it
is always a good practice to initialise weights using ‘He initialisation” or “Xavier’s Initialisation”.

Tensorflow

At a high level, TensorFlow is a Python library that allows users to express arbitrary computation as
a graph of data flows. Nodes in this graph represent mathem - atical operat ions, whereas edges
represent data that is commun - icated from one node to another. Data in TensorFlow are repres - ented

Tensorflow (cont)

> as tensors, which are multidimensional arrays. Although this framework for thinking about
computation is valuable in many different fields, TensorFlow is primarily used for deep learning in
practice and research

How to write a code to start session for the training?

with tf.Session() as sess:

Image Segmentation

Image segmentation is a further extension of object detection in which we mark the presence of an object through pixel-wise masks generated for each object in the image. This technique is more granular than bounding box generation because this can helps us in determining the shape of each object present in the image. This granularity helps us in various fields such as medical image processing, satellite imaging, etc. There are many image segmentation approaches proposed recently. One of the most popular is Mask R-CNN

Instance Segmen tation: Identi - fying the boundaries of the object and label their pixel with different colors.

Semantic Segmen tation: Labeling each pixel in the image (including backgr ound) with different colors based on their category class or class label.

Sponsored by CrosswordCheats.com
Learn to solve cryptic crosswords!
http://crosswordcheats.com