*) Expert systems.

(example problem)

bi classification → being able to o/p. whether an image has text in it or not.

typical yes or no problem

↓

primary goal : being able to train a machine to take/make decisions (smart decisions)

↓

Human decision making process

↓

identification of important key features related to the problem.

← (dengue / disease diagnosis)

symptoms for the specific problem.

(i) learning the dependence of different features on o/p.
(function approximation)

↑ approximating coefficients as well

(ii) representation learning
(feature extraction)

↓

Based on the past experience of the comb. of values of the key features identified by the domain expert, classification is done by the expert.

← another example could be making decision for lbw.

↓

interested in allowing machines to take these decisions

↓

i) encoding the data to numeric form
(allows it to be parsed by the m/c)

↓

semantics of rule making : i) features
ii) rules.
(coming up with certain combination of features to make decisions)

*) outsourcing this decision making responsibility to a m/c.

↓

one method can be a combination of if / else conditionals, that act as the rules in the expert's head, used by the system to make decisions.

(typical expert system)

↓

have been used for pretty high level state of the art applications / use-cases

*) limitations: i) too many factors impacting the problem, then defining all permutations & combinations for an expert system might be very unfeasible. and cumbersome.

↓

(rules definition for dependence of o/p on different factors)

ii) The rules might be too ~~[illegible]~~ complex or inexpressible.

iii) The knowledge might not be enough. No matter how knowledgeable the expert, there is no guarantee that they would be able to figure the correct rule for appropriate classifications or predictions.

*) Machine learning

↓

given certain features / factors, the decision making process can be said to a function of these features or factors.

↓

Family of functions

in expert systems : a human was establishing this function.

↓

being able to eliminate human from process of figuring out or approximating the function.

↓

And allowing the machine, given the data, to approximate this decision making function, along with different parameters / coefficients relating these features to the decision making function, which best relate the given i/p .dataset to o/ps.

← being able to make sense out of the data to make better predictions.

↓

algorithms to perform this approximation

↓

DL has one family of algorithms which allow us to make this approximation

•) Reasons for inc. in ML relevance : i) Abundant data

↓

multi-modal (forms) (different types of data) → text, audio, video, numbers, image

multi-lingual (different languages)

ii) Democratized model + learning algorithms
↓
availability of APIs or codes for experimentation + use
(opensource)

iii) Relatively fast and cheap cloud computing
↓
better hardware
(gpus, tpus)

*) Different roles in ML world : i) collect, curate data and define tasks
(i/p → o/p)
↓
being able to recognize + realise the possible tasks that can be completed using the available data, and being realistic about them.

ii) ML Engineering : figuring the right algorithm to be applied on the problem.

iii) ML Researchers : figuring out better and faster methods of establishing the required goals or applications.
(better DL model,
better method of training a graphical model)

Scanned by CamScanner

*) six elements (6 Jars) of ML

(framework to categorise, any activity related to ML to one of these Jars)

↓

being able to categorise and filter the jargon cloud to proper categories, for better and efficient learning and understanding

↓

I) Data → a) structured data (tables, feature extracted data)

b) text

c) image

d) video

e) audio.

can be high dimensional (R^n)

supervision

↑

being provided with example labels

↓

For supervised based learning, it is important to have mappings b/w x and y, and the data should be in machine readable format.

that too suitable to the problem at hand

↓

will differ based on classification, regression (exact bounding box coordinates)
(0/1)

apt encodings. (text, video, image, audio, etc.)

↓

Data curation: i) open resources/sites: google ai
data.gov.in
uci

ii) platforms for annotation:
amazon mechanical turk
dataturks
figure eight

crowdsourcing the annotation process

↓

pretty powerful.

iii) self curation / annotation

extraction and all

wikidata
(words in
one language to
another
language)

2) Tasks : possible tasks that can be performed

crisp defn. of what our i/p is + what our corresponding o/p is.

given a certain form of data

i) unstructured data → structured data.

(given the unstructured data (specifications about the product) + being able to populate the database of the company using the given unstructured data)

ii) reviews + specification → generation of FAQs

(key questions / problems highlighted in the reviews, being able to generate some. from reviews itself or from the provided specification)

iii) given ii), being able to answer variable / new. customer questions

Battery draining out

*) object / face recognition → image tagging.

(coordinates → label for the corresponding bounding box)

*) activity recognition

*) location recognition / identification

*) recommendations.

tasks → supervised (data along with its corresponding label)

↓

classification

(segregating or identifying the correct label for the corresponding i/p)

→ bi classification

→ multi-class classification

← most appropriate label out of the ones possible

↓

regression → dealing with real values, for eg. exact coordinates of the bounding box for the objects in the image., or predicting stock prices, house prices, etc.

unsupervised → finding / identifying patterns in the image dataset and accordingly creating clusters for the same and organising them into the apt. group.

generation → generating x's similar to the x's in the main dataset.

↓

GANs. (Generative Adversial Networks)

↓

ex. image generation
tweet generation

3) Models : Mathematical formulation of a task

$y = f(x) \rightarrow$ true

$\hat{y} = f(x) \rightarrow$ approximate

↓

The tried out functions on the past data are the models which try to formulate / model the relationship between different x & y. → (approximate)

DL → NN family of functions

why not using a complex model always is a good idea

↓

this may lead to overfitting and over-specification of the problem, which should be avoided as much as possible.

4) Loss function, acts as a metric of being able to judge, which model is better suited to the data at hand, out of the possible alternatives.

↓

by analysing how far away the predicted values using the given model are from the real values. → (squared error)

↓

i.e. sum of the diff. of ~~values~~ true & predicted val of the dataset.

↓

we need to negate the effect of signs, and somehow operate on abs. values somehow.

↓

∴ modulus of the diff. (abs diff) (magnitude)

↓

and then we take the square (due to another calculus based reason)

Scanned by CamScanner

∴ squared error is taken for a model, and that can act as a metric on how well a model is an approximation of the true values or fn.

↓

and this can thus be used to choose the most appropriate approximation.

↓.

∴ loss fn allows us to see how good or bad our model & its corresponding parameter/coefficients approximations are.

↓

i) square error loss
ii) cross entropy loss
iii) KL Divergence

5) Learning algorithm : identification of the parameters of the model

↓

can be thought of as a search problem, intially.

↓

mechanism/algorithm that allows the efficient approximation of the parameters or coefficients such that the loss function is minimized.

↓

becomes an optimization problem, i.e. identification of parameters such that the loss fn. gets minimized.

↓

i) gradient descent ++ (and its variants)
ii) adagrad
iii) RMSprop
iv) Adam

data collection
task selection
model
loss fn.

} human oriented

learning algo. → machine oriented.

↓

since this is m/c oriented, we can take certain liberty, to allow in one way to ask the m/c to approximate the most suitable fn. and its parameters, by providing an exhaustive list of possible fnc.

6) Evaluation of a ML model

↓

i) accuracy : abs. accuracy which takes into account the ratio of correct predictions and total predictions.

ii) top-3 accuracy : allow for 3 predictions for the same i/p. if any one of the 3 is correct, it is counted as correct, otherwise, it is not.

(for example search engines if the needed id / website / link is among the top 3 or 5, the user would mostly be satisfied)

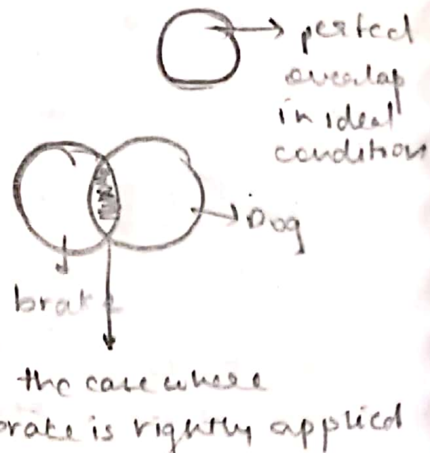
↓

How is this different from loss ?

↓

accuracy is a more & better interpretable metric as compared to loss.

considering the problem of an autonomous traversal vehicle, which needs to keep moving till no obstacle is encountered.



↓

iii) precision-recall : precision : instance when correct action/prediction was made

(another metric which is used heavily)

recall : total actions, taken

we work on optimization of loss fn. for training data, and that will not be done on testing data, whereas evaluation will be mainly done on testing data (better evaluation)

↓

data not encountered by the algorithm.

imagenet

pascal 2

Abundance ← Data

↓

Task

↓

Democratized ← Model

↓

Loss

↓

improvised

Learning

↓

Standardised ← Evaluation

Linear Algebra

Probability

calculus.

foundations.