CS 4375: - Introduction to ML

. Basic knowledge of Math & Prob

. Python

**Day1: ML Introduction**

Machine Learning:

You have probably written programs (code) in the past.

• You had to program every possible situation: Think of "if – then – else" statements.

• What if machines could learn by themselves from their environment or data.

**What is ML**

. to teach a machine to learn to do a particular “task”

. How?

. so far, explicit instructions

Enter num 1:

Enter num 2:

Answer = num1/num2

I check if num2 == 0;

If (mouseMove) do this;

Else if (mouseClick) do that;

. In this class? Get the machines to learn by themselves using training data

. What do I need to do? Training data supplier

. MONITOR the learning

. Machine will try to find pattern in the data and use that for learning.

. What type of tasks?

. Teach the machine to predict something.

House price prediction.

. online sales

. gas prices

. holiday sales

. GPA

. Weather/ rainfall

.

There will be certain uncertainty

Kaggle.com: Coding competition

Machine learning:

. From the instruction or from data

. Learning from data not from instruction.

Machine Learning

• Machine learning is a method of data analysis that automates analytical model building.

• Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

• Sounds too good to be true? Well, it's already being widely used.

Let's look at some examples

.

Machine learn from pattern -> model

. training data

. predict a value (house price, spam/no-spam).

. Learn a model from Training Data.

. Apply the model on b=new Data (Test/unseen Data).

Model

* Pattern identifier
* Decision tree
* Neural nest (deep learning)
* …

4 free day.

At most 2 free days for any given assignmen.

Assignments:

Team of max 2 students

Project:

Team of max 4 students

Z – estimated : Zillow estimate price of the house

Name:

ML application in your daily life:

. Amazon recommendation

. GoodReads

. Spotify

. Voice Recog

. Sport App

. Instagram Reels

. Twitter

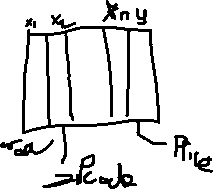
.

How did you spend your summer:

. Basics Machine Learning (X, Y -> target, what to learn)

. Prediction: no free lunch theorem:

. Known ->(risk) unknown



. Specific ->(lose certainty) general

. Zillow – Z-estimate:

1. Supervised “What” to learn (X, Y -> target, what to learn)

. Learn to predict house prices

. Learn to predict spam or no- spam

. Weather (temp)

1. Unsupervised

. Data mining

. Tell me something new

E.g.: . outliers

. group (clusters)

. association rules bread 90%-> butter

. Move x -> move Y

….

**Prediction function (S)**

Y= W0 + W1X1 + .. + WnXn

**Recommendation (Uns)**

X = (x1,x2,…,xn)

1. Reinforcement Learning (No data)

. Agent -> Rules -> reward punish

Diagram

Description automatically generated.max reward

Diagram, engineering drawing

Description automatically generated

…

**Predicting a target variable based on feature of data**

1. Discrete, categorial, class

. [0,1] ; [A B C D]; [yes no]; [sleep Eat Play]; [Dog Cat]; [Pass fail]; [Spam no-spam]; [chin, amer, Indian, cam]

. Classification

1. Target is a Real number

. 0; 0,00001; 1,909383

. Regression

. Predicting House price, GPA, Weather

Both are Supervised.

….

Regression problem -> Classification (yes)

Score in the class Thresholding

90.38 >70 Pass (comment)

67.16 <=70 Fail

90<=S<=100 ->

100 <= S <= 120 ->

120<S ->

….

Chapter 2:

Motivating Example:

. Classification

Chart, scatter chart

Description automatically generated

. Regression

*Deep Learning more about (preticting)*

.

**Components of ML**

What do you for a ML project? (3 components)

1 • A well-defined task - what are you trying to teach to the machine?

2• Training data that will be used by the machine – also called training experience.

3• A way to check the performance of the machine – called error metric.

• A smart machine learns from the past errors and improves its performance

Performance measure

Text

Description automatically generated

h-> current hyp

y -> true value

MSE = 1 \* (h-y)2 / N

“A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.” -- Tom Mitchell, Carnegie Mellon University

P ^ - Error giam -> more Exp

Graphical user interface, diagram, text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Supervised Learning – Inductive

• Inductive: Tries to discover general concepts from a limited set of training examples. => Generalization

- Based on search of similar characteristics in different classes of examples.

- e.g. Given labeled examples, you find features that are common within each class

• **Inductive: goes from specific to general**

- tries to obtain new knowledge.

- new data points may force you to change old hypothesis

**Inductive vs Deductive**

Deductive: uses given premises and logical arguments to infer conclusions. General -> specific

• tries to obtain knowledge that is implicit in original knowledge.

• Classic example: \* All men are mortal. (major premise) \* Socrates is a man. (minor premise) \* Socrates is mortal. (conclusion)

Learning inductive :

A picture containing diagram

Description automatically generated

Diagram

Description automatically generated

X-> image (data) feature

Y -> label

‘f’ –

How to find the best feature ?

Diagram

Description automatically generated

No right or wrong , have to decide.

Depend

**Inductive Learning.**

Training :

. Training data (Sample Data)

.- Create a model

Test (unseen).

* Accuracy
* Overfitting

. Model works good on training, and it generalizes poorly

. over-complex model tends to overfit.

. Limit model complexity

. Relation between training and population parameters

Sample 🡨--> population (true)

N =100 150 mn

. Sounds great, but.

. Hoeffding’s inequality

. Learning Process

. Mean square eroll

(y1 – h1)2

08/31/2022

Minimize the error

. convex function

Neural network

min error function

Parameters:

weights joining each layer

Data ->

1. Gradient Descent

4 lines

1. Real world

* Scikit learn

. way of minimizing a convex function

. error/ cost

….

Scikit learn

* OLS: analysict

…..

Most of the time should be spent on data pre-processing visuallation

Trees 🡪 Randoms forest

..> boosting adaboost; xGB (Extreme radient boosting)

Bad:

* Overfit
* Early terminating:

Setting max-depth

If IG of split is not significant, don’t split.

. Tree pruning: let the tree grown without any stopping. When the tree is fully grown: prune the leaves.