## [CAP4611-21Spring](https://webcourses.ucf.edu/courses/1369384/calendar_events/2158980)

# Variant 1

# Day 1 (Tue, Jan 12):

Required Text: [Amazon Link](https://www.amazon.com/Fundamentals-Machine-Learning-Predictive-Analytics/dp/0262044692/ref=pd_ybh_a_3?_encoding=UTF8&psc=1&refRID=9KANHYJ9YCGR072ZT7FW) | Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies 2nd Edition

Recommended Practice (For Later): [Kaggle](https://www.kaggle.com/)

**Reinforced Learning:**

* Toss the system some input, let the system run with it.
* Once finished we let the system know whether the output was good or bad.
* Model will adjust, the system will try again until it frequently runs successfully.

**Unsupervised Learning:**

* Tons of data, no labels
* Does the data mean something?
* Trying to find something interesting in the data. Are there relationships? Is there a way to reduce the amount of data you’re working with(?)

The Algorithm is a mathematical model that you’re trying to figure out, training is trying to figure out what the parameters are. Parameters are adjusted during the training process. Training is adjusting those parameters by some sort of mathematical search method.

**Decision Trees:**

* Basically just binary trees.
* How many layers deep should it go?
* Why is it a decision tree? There are decisions related to the edges(?), e.g., x > 7, x < 5.

AI used to be codifying(?) a set of rules in order to make something work as intended, e.g., natural language processing.

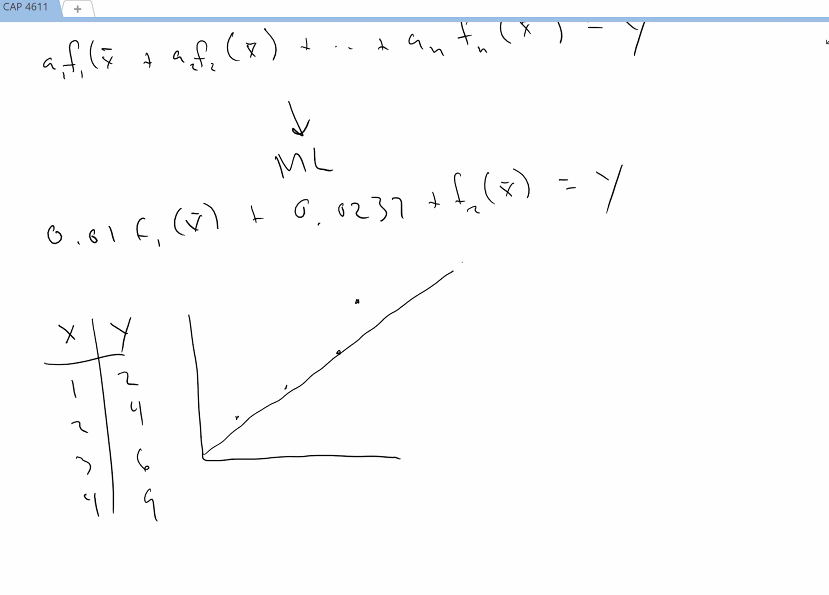
Study supervised learning

Lecture 1: Machine Learning (General Idea)

-> Looking at general linear regression models.

-> Supervised machine learning vs unsupervised machine learning vs Reinforcement learning

Supervised machine learning:



X values:

1

2

3

4

Y Values:

2

4

6

9

# Variant 2

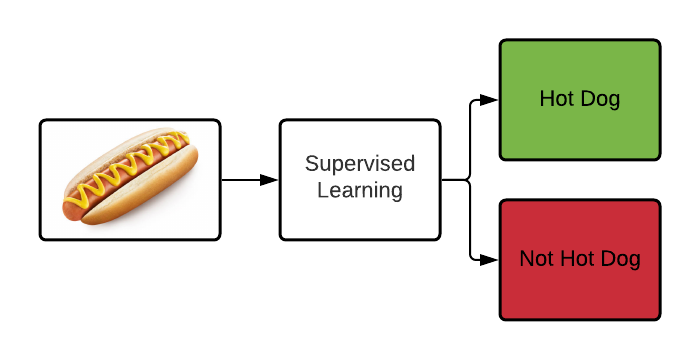
So you may have in 3 dimensions, and some sort of surface that wanders through in 3 dimensions that classifieds everything on one side and everything else on the other side

The notion of drawing a line through points is linear regression, we will go into more detail in this class….

So that's great, its useful , but when we are talking about machine learning, we are talking about 3 subgroups, supervised, unsupervised, and reinforced learning

Supervised machine learning:

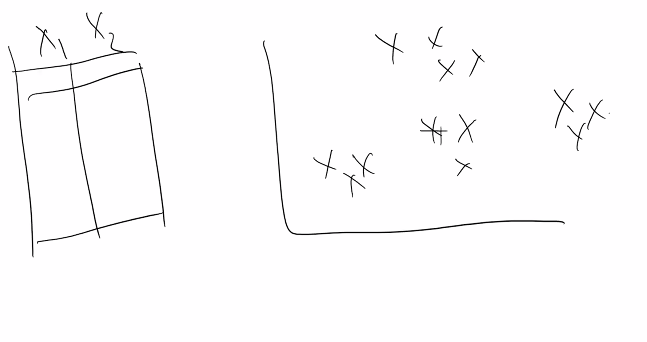
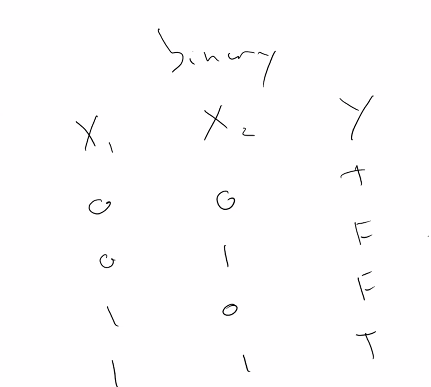
* Basically, inside this data, I know what output is produced by those inputs. So if you know about shopping, if I know a certain set of products, or spend a certain amount of money on brushes or conditioner, I can look at what things they have bought.
* House prices, if you look at features of a house sqr. Footage, you can look at how much a house is worth
* Basically, looking at a whole bunch of input data and receiving some sort of output data.
* So, if we can produce an output generated by data that is not available to the unit



Reinforcement learning:

* Not covered because of dependencies
* If you think about playing games, starcraft, UAVs etc etc, that is basically reinforcement learning. YOu just let the system do something and give it a bunch of output, if you can tell whether the system is good or bad, the model needs to be adjusted.
* As the system plays the game over and over… The system will eventually be able to play the game (think about alpha GO)
* There are all sorts of interesting videogame ideas
* OpenGM?
* Deep reinforcement learning - if you feed them a screenshot of the game over and over and over, it will learn to play the game
* The other type of learning, so we have supervised learning, we have unsupervised learning, that goes into machine learning...

Unsupervised learning:

* What you do is you have a bunch of data
* The only data you have is input, no output LABELS
* What we want to know is “I have this data, is there any sort of meaning?”
* Example
  + F(X1, X2)
  + If we want to visualize this data, it would look like this
  + 
  + Are there any clusters of data?....
* When we are doing machine learning and data science we are doing it with an absurd amount of dimensions.
* If you think about an image, it could be 100x100 px. How many pixels are there? 10000
* So an image that is 100x100 has 10,000 different inputs.
* Now imagine you want to do machine learning on that image. How many inputs is that? A bunch
* You will learn that as your input space grows, it takes a lot of computation. But even more importantly, the more inputs you have the more features, the more data you have.
* This is called the curse of dimensionality(?)
* If we think about binary data, lets say we want to deal with data that has two inputs
* If this is binary, what values could x1 and x2 have? True or False
* 
* If we have 3 inputs, we will need 8 things
* If we have 4 inputs, we need 16
* If we need k values, we need 2^k values.
* As we add more features, the amount of examples grows exponentially.
* …..
* Columns are features, rows are (???)

Let's say we have y = ax + b

Our output is y

A is a parameter

B is a parameter (aka bias)

So the whole notion of training an algorithm, the algorithm is the mathematical model you are trying to figure out. Figuring out the (?)

Now there is another thing called hyperparameters

The parameters are adjusted during the “Training process”

Then there are hyperparameters, which allow you to try different models. Things that require a different structure.

When we talk about building a binary trees…

How many layers should it go down?

We may have one model that goes 3 layers deep, another model that goes 4 layers deep, etc etc.

There are many different conditions… but we’ll talk about it later

What I want you to know today is that there is:

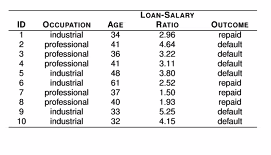
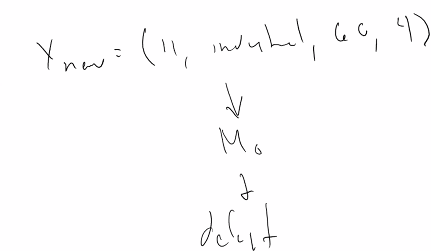
Supervised learning - you know the answer, you train the system so that it can find the answer given input

Unsupervised learning - We don’t know the answer and we are trying to figure out something interesting about the data.

That's great… but why would we need machine learning?

* The way AI used to be approached, is with this notion of rules
* We gotta figure out a way to “codify” the language to make (?) work

So lets… say we have a dataset:

* 
* We would like to figure out what the outcome is based on this data. So who can think of a way we might be able to predict the outcome given this data?
* How do we know what should be repaid and what is default?
  + What people have said
    - Industry vs professional
    - Loan salary ratio
* We need to figure out some function that gives us the necessary output.
* There is some notion of thinking about using occupation
* We can see industrial → repaid
* Professional ---> default
* If we just look at the “industrial” feature we cannot definitively tell.
* Looking at the ratio column,
  + If we took the loan ratio and sort the data on it, what we will find is
  + We get 2s 3s 4s…
  + Large values at the top small values at the bottom
* What we will be able to see is that, if the loan salary ratio is > 3, then the outcome == default
* Otherwise, the outcome == repaid
* This is a model for this data. (referring to a conditional statement)
* Some formula
* 
* How do we know that the data is right?
* ….
* On small datasets, we can figure out some rules, but what happens if we have an even larger dataset?

Question:

Would it be logical for a credit card company to have some sort of logical data like that?

A: yes, they have invested in that technology. If you swipe your card in Mexico but all your previous things are from BP on university, something is fishy. When you go for a loan, their models look at your things too. If we

….

* If we look at the dataset, can we have a set of rules that cleanly splits the outcome into repaid and default? Yes but it will be very difficult. This is where ML comes in.
* As we get more data, it will become infeasible and then eventually impossible for us to do.
* This is exactly why everybody wants your data. They just want to scoop up tons and tons of data to compare profiles to see patterns and correlations.
* Fields: Marketing, streaming, Social Media, etc etc
* Think about all the free data you give to Amazon or Facebook.
* If they track everything you buy and you stop buying from a certain store, they could send you a coupon to come back.
* Issues with medical data: We have privacy laws to keep it private

Tor browser - great for hiding where you are, basically the web browser “skips” and “hops” around to hide your connection

(Suggesting you to use Tor)

If you don’t use Tor, get some anti-tracking plugins or add-ons

Question:

We are probably going to only talk about things we (?)

You gotta figure out what inputs you need to collect. Once we get those inputs, how do we compensate?

….?...

So in those cases, we can reduce images down and use that as the input to our algorithm

* Out of time
* In terms of languages, Python will be what you’ll be using
* For statistical work R is better
* There is nothing wrong with using R to examine your data and move into Python

Question:

I saw you suggested R for data science (book)

Authoer makes it available for free

For Python what would you use?

There is so much stuff out there for python, just use anything

Kaggle has some tutorials

We’ll get more into it thursday.

I would say, the easiest think would be to use linkedin data camp (?)

The python tutorial website is pretty good

...Plenty of sources for python out there…

“So that being said, end of recording”

So for GPU stuff that is mainly used for Neural networks

But there will … (?)

Uh so yeah, the reason GPUs is used for machine learning, they are used for matrix math, and Neural networks are almost all matrix math.

It all comes down to speed, the reason Nvidia stuff is so expensive, is because of how expensive it is.

Question:

…?

Things in linkedin learning are a lot simpler, we don’t really know much about a topic, linkedin gives you a general knowledge about a subject, then you gotta look at a textbook, once you are done with taht you gotta look at research.

Blogs are ok, some explanations are better than others, but if you really want a good explanation of the mechanics behind something, you gotta look at a textbook, research etc etc

If you only limit your learning to machine blogs etc etc, you’re limiting your knowledge (?)

Best to Look at the actual documentation for an API or anything at all (this is important)

Also, consider running test functions to see if code within a new library actually does what you think it should do. (unit testing)

Javascript, Python, and many other languages & frameworks have their own testing languages

Considering main functions with unit tests,

Question:

Would you consider talking about the machine learning field later down the line?

I can tell you right now:

The easiest way, because there are so many websites and compenttion websites, compete, learn a bunch of stuff in them, and once you have that experience, you can go into a job interview and actually talk about it.

Question:

I heard its easier to go the data science method. Will it be possible for undergrads to enter the data science field?

a : Titles are useless, we have this title, that title, so many different titles, basically the all do the same stuff.

If you find a company that knows the difference, you can know that some jobs ???

You are not usually using the (??) writing machine learning from scratch(?)

I’ll see you all on thursday:

Hopefully, when you start the competitions, you want to play around with a bunch of datasets and get an intuitive (abstract) idea of how to extract and mange this data.

UCF has an AI group and a Robotics club that cover these things.

Some question:

So there is a big difference between teh answer being a 0 and a 1 or some integer

Lucas question:

“After this class are we gonna be able to start reading the literature and getting up to date on the new things in ML or will we need to go through the Bishop book first before we can understand the papers etc?”

Gotta read your papers yourself, do the math, find out

Benefits of advanced degree:

Paycheck, its a lot easier to get your foot in the door with a adv. Degree in machine learning. To be honest, there is so many resources out there where you could become a master yourself, but HR loves paperwork.

…

…

…

Alright, you all have a good week.

Q: do graduate students go over measure theory?

Eventually, but you don’t need it to learn it to learn Machine learning

f.closeZoom()

I need a a friend for pair programming:

notjohn#1626

'yer Pappy speakin

#9709