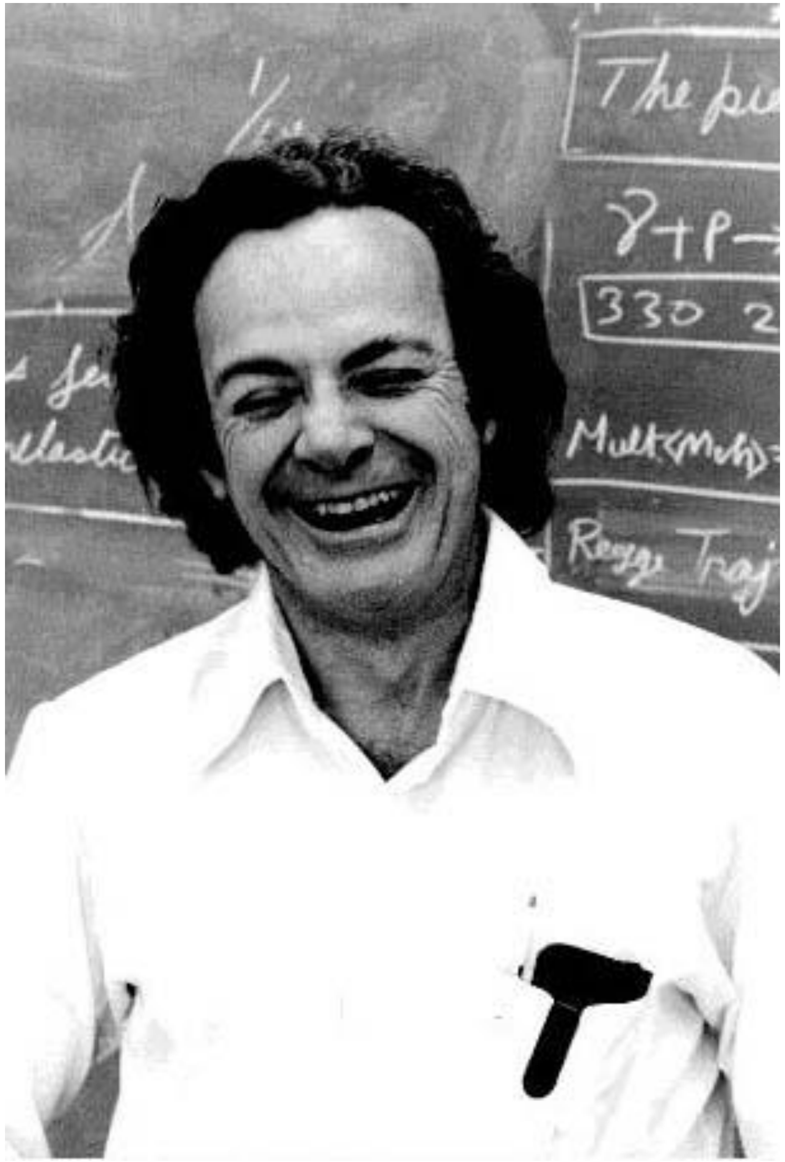
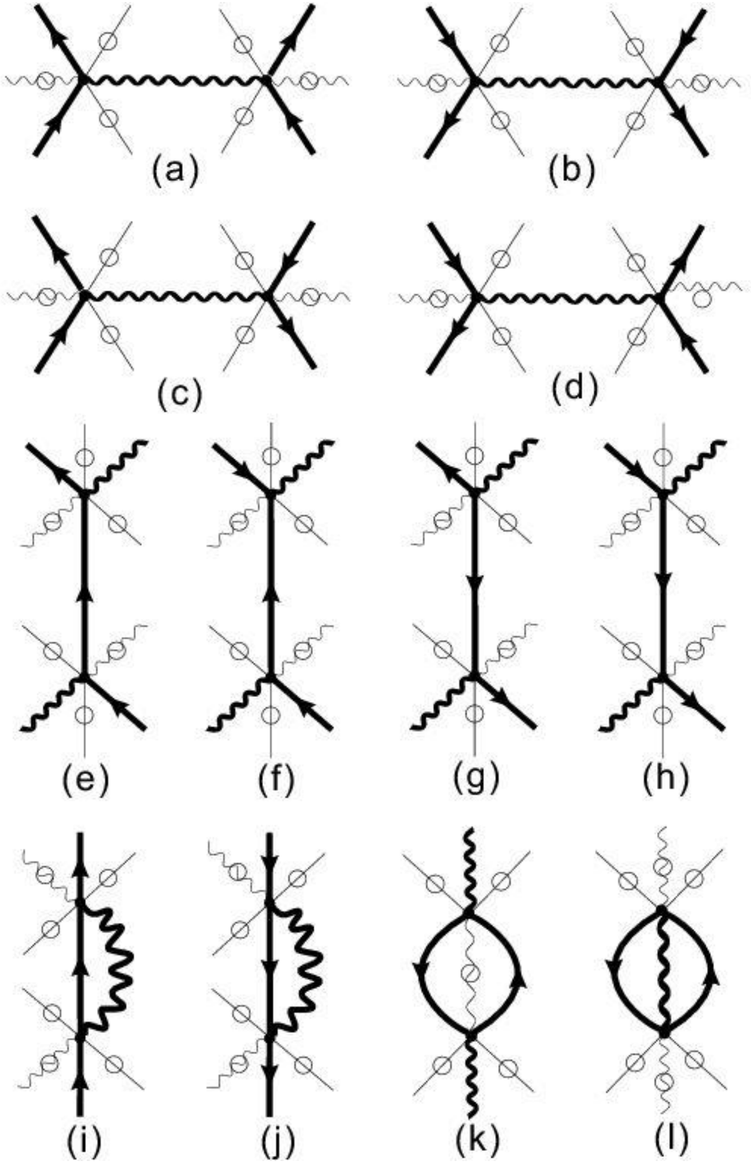
OVERVIEW

MGMT E-5072, Fall 2018

Data Literacy in the Age of Machine Learning





"To understand *how* subtraction works – as long as you don't have to actually carry it out – is really not so difficult. That's my position: I'm going to explain to you what physicists are *doing* when they are predicting how Nature will behave, but I'm not going to teach you any tricks so you can do it *efficiently*. … It takes seven years – four undergraduate and three graduate – to train our physics students to do that in a tricky, efficient way. …By explaining quantum electrodynamics to you in terms of what we are *really doing*, I hope you will be able to understand it better than do some of our students!" (Feynman, QED p.12)

When you work with (and depend upon) others who know things you don’t and vice versa, you have to be able to work productively with them.

Together you can become more than just the sum of your parts.

"More broadly, companies must have two types of people to unleash the potential of machine learning. 'Quants' are schooled in its language and methods. 'Translators' can bridge the disciplines of data, machine learning, and decision making by reframing the quants' complex results as actionable insights that generalist managers can execute."

An executive's guide to machine learning,
McKinsey Quarterly, June 2015

It's not just about reframing the results but about framing the problem correctly in the first place.

A few AI applications today

What makes these possible? Computational power, the availability of data, and …

A LOT OF NUMBER CRUNCHING

- BUSINESS INTELLIGENCE
- IOT PREDICTIVE MAINTENANCE
- SEARCH RECOMMENDATIONS
- FORECASTING MODELS

VISION

- AUTO TECH AND DRONE COLLISION AVOIDANCE
- E-COMMERCE SEARCH
- PICK AND PLACE ROBOTS
- HEALTHCARE DIAGNOSTICS

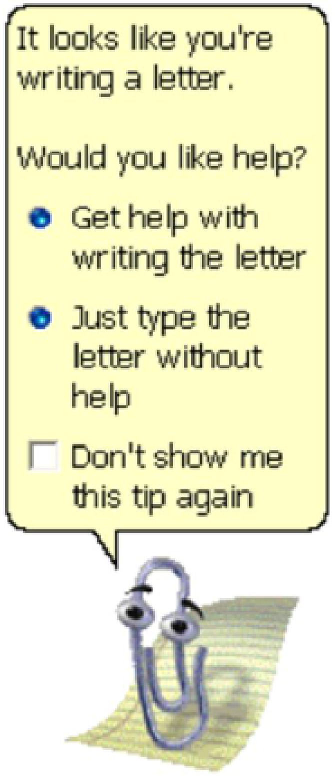
LANGUAGE PROCESSING

- CHATBOTS
- NEWS & MEDIA CONTENT CREATION
- SMART HOME VOICE INTERFACES
- TEXT ANALYTICS

Source: www.cbinsights.com

FROM CLIPPY TO ALEXA…

… and learning without explicitly learning a set of rules. Think about how you learn to hit a tennis backhand or how you learn to recognize faces in photos. It's not by writing down an explicit set of rules for doing these things.



"Alexa, ask Uber to request a ride."

"Alexa, ask Fitbit how I slept last night.

"Alexa, ask the bartender, what's in a Tom Collins?"

"Alexa, tell Tide I have a juice stain."

Source: www.cbinsights.com

DEFINITIONS OF MACHINE LEARNING

"[Machine learning is the] field of study that gives computers the ability to learn without being explicitly programmed."

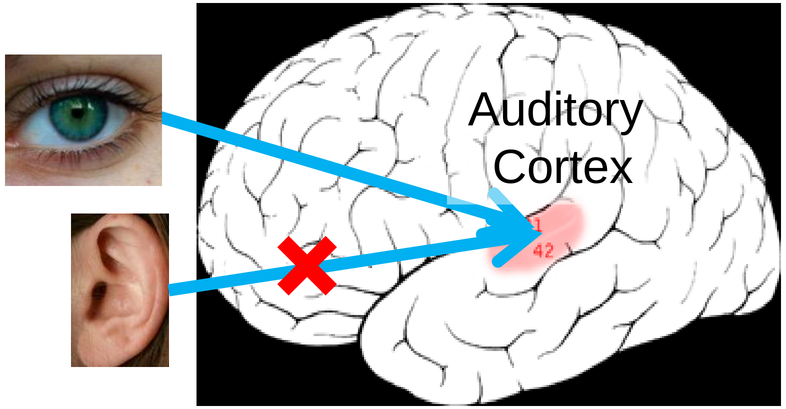
-- Arthur Samuels *(1959)*

"A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."

-- Tom Mitchell, *Machine Learning (1997)*

THE ONE-LEARNING-ALGORITHM HYPOTHESIS

Sensory cortex learns to see.



Auditory cortex learns to see.

When you give 2 different brain areas the same inputs, you get the same outputs.

+

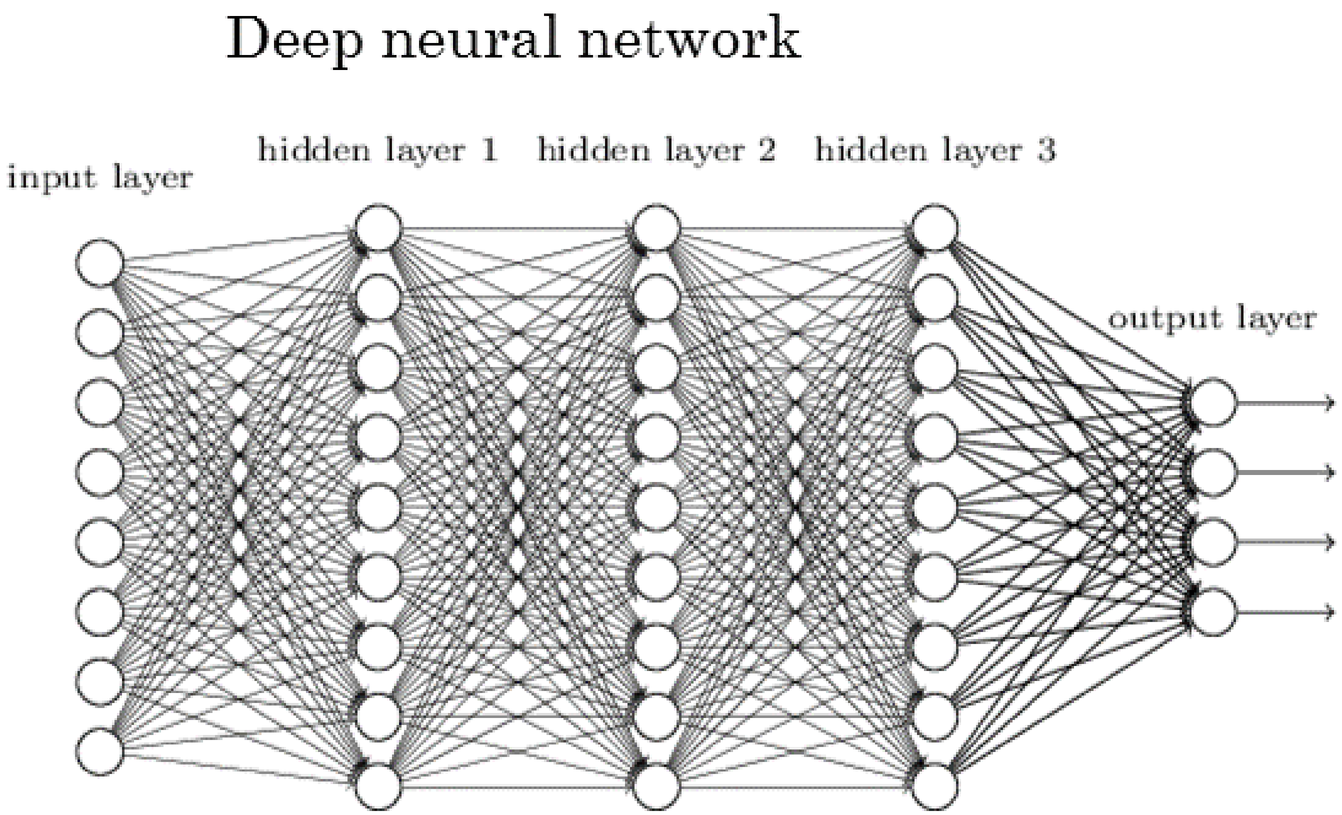
When you give the same brain area 2 different inputs, you get 2 different outputs.

=

Conclusion/Hypothesis: No matter the brain area, it's doing the same thing – i.e., running the same algorithm.

Source: Andrew Ng; BrainPort; Martinez et al; Roe et al.

DEEP LEARNING



Source: GoogleBrain

Universality Theorem: A three-layer neural network can calculate *any* function (to a close approximation).

CAUTION: That a function can be *calculated* doesn't mean it can be *learned*.

Check out Michael Nielsen's online book for more information: http://neuralnetworksanddeeplearning.com/chap4.html

Source: McKinsey Quarterly, July 2016

THE SINGULARITY

“Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.” (I.J. Good, 1965)

“People sometimes ask how quickly I think we will get there, and my honest answer is I don’t know. We could get there in 3 years or in 30 years. **But I do believe that it will happen in this century.**”

Marek Rosa

CEO, GoodAI

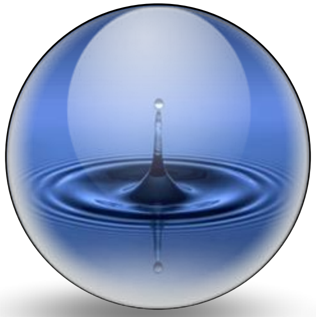
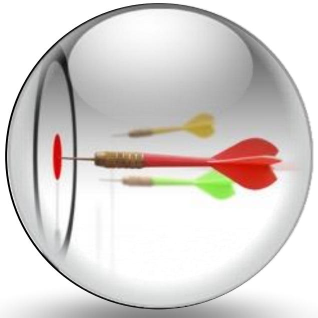
"Despite these astonishing advances, we are a long way from machines that are as intelligent as humans—or even rats. So far, we've seen only 5% of what AI can do."

Yann LeCun

Director of research, Facebook

Source: www.cbinsights.com

THE COMPONENTS OF INTELLIGENCE

	Perception	Learning	Reasoning	Abstraction
Rules–Based Systems	No	No	Yes	No
Machine Learning	Yes	Yes	No	No
Deep Learning	Yes	Yes	No	Yes
Contextual Adaptation	Yes	Yes	Yes	Yes

Source: DARPA

WHAT WE'LL COVER IN THE COURSE

- *Part 1: The Mechanics of Prediction*. In Part 1 we'll dive right into machine learning, unpacking the key concepts. We'll apply these concepts to make *predictions* from real datasets. We'll cover the basic techniques of machine learning – regression and logistic regression – and get a feel for the practical things that data scientists do. We'll round out this part by taking a look at more advanced machine learning techniques.
- *Part 2: The Science of Machine Learning*. In Part 2 we'll learn to systematically evaluate the performance of machine learning models. We'll understand how to define performance and measure it. We'll use this knowledge to not only build the right machine learning models but build them right.
- *Part 3: The Art of Machine Learning*. In Part 3 we'll tackle the art of machine learning – how to get the most predictive bang for our data buck. In other words, we'll learn about how to make the most out of the data we have.
- *Part 4: Select Topics in Machine Learning*. Finally, in Part 4 we'll cover select topics in machine learning: segmenting customers, spotting fraud, detecting spam, and building a machine learning system.

PREREQUISITES

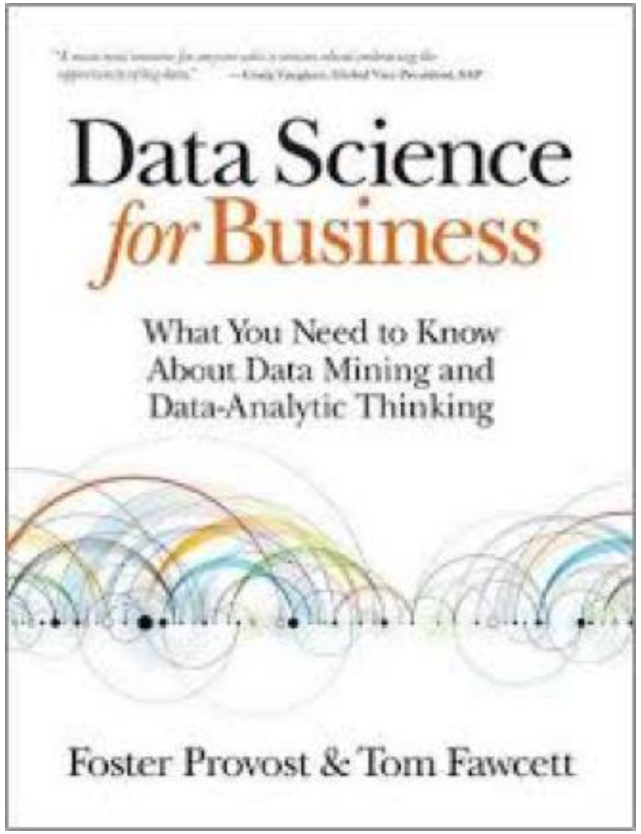
- We do not use any advanced mathematics in this course. If you've taken the SAT or the GRE you've already come across math that is much more advanced than anything you will need for this course.
- Alternatively, if you're comfortable working with spreadsheets (nothing fancy, just basic formulas and manipulations like sorting rows), you will be comfortable with all of the mathematics used in this course.
- Hands-on learning is encouraged using the Orange data science platform (https://orange.biolab.si/) – a visual way to solve machine learning problems without programming.
- For those with some programming knowledge of Python, we provide Jupyter notebooks that can be used to build, run, and experiment with machine learning models.
- NOTE: Python knowledge is NOT a prerequisite for the course. The course assignments (homework, group presentation, and the final exam) do NOT require any Python programming.

WHAT IS *NOT* COVERED IN THIS COURSE

- We will not cover the details of machine learning algorithms – how they are designed or how to make them efficient.
- We will not cover the theoretical underpinnings of machine learning – probability, statistics, statistical inference, numerical computation, etc..
- We will not cover computer science topics such as data structures, algorithm design, search, etc..
- We will not cover deep learning and associated platforms like TensorFlow, Keras, Theano, etc.
- We will not cover data architecture/engineering for implementing machine learning systems at scale.

This is not a course for students seeking to learn the theoretical and mathematical underpinnings of machine learning.

OPTIONAL TEXTBOOK



- Nice alternative view – useful when you're learning something new.

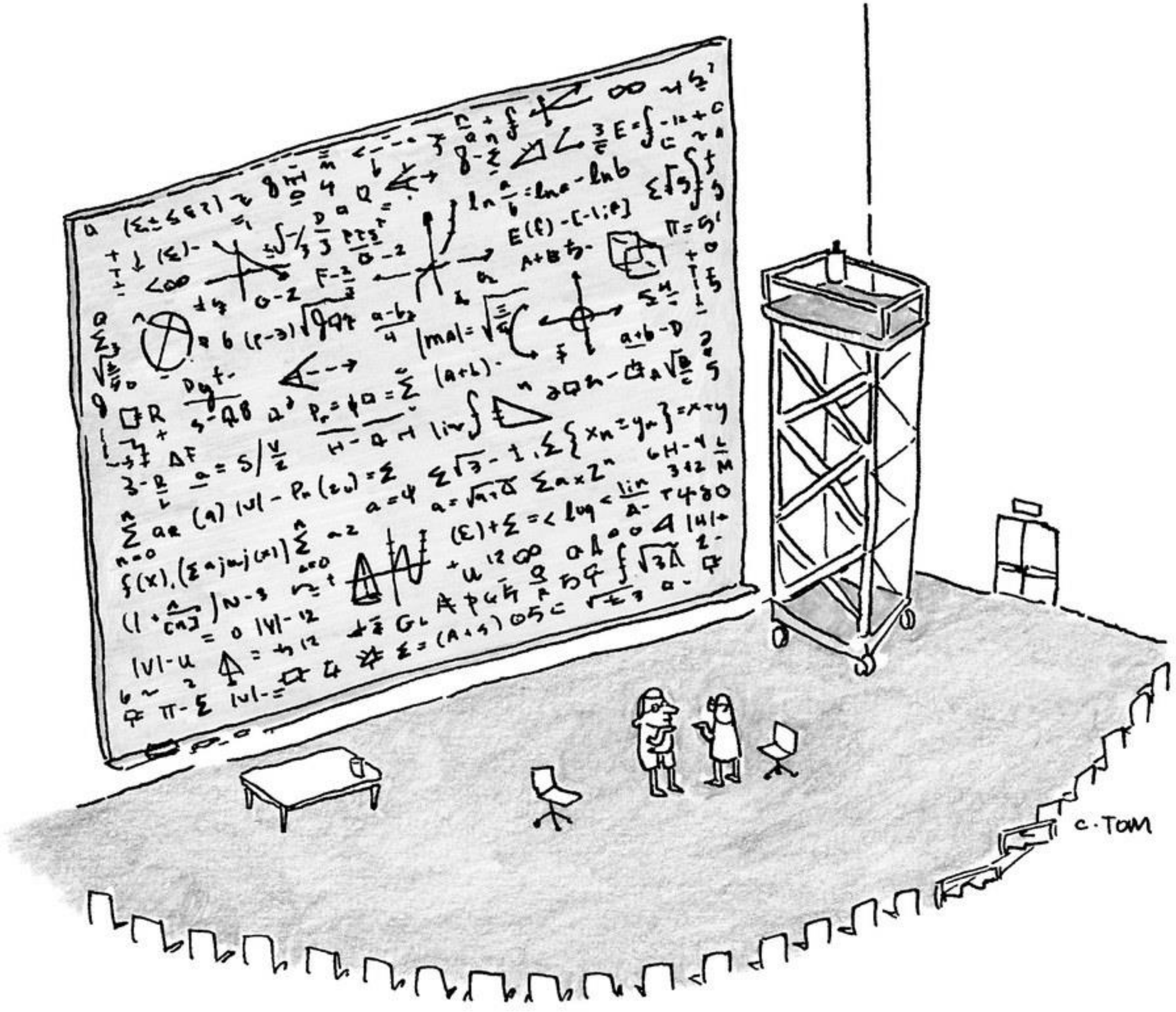
- Strong on the business aspects of data science – chapters 1, 2, 11, 13, and 14.

- Additional book suggestions:
  - *DataSmart* by John Foreman
  - *AIQ* by Nick Polson and James Scott
  - *The Master Algorithm* by Pedro Domingos

- Lots of useful blogs/websites:
  - Adam Geitgey (*Machine Learning is Fun*)
  - Jason Brownlee (*Machine Learning Mastery*)

BY THE END OF THIS COURSE YOU'LL BE ABLE TO…

- List the types of problems that can be solved using machine learning.
- Understand the seven key steps to solving any machine learning problem.
- Apply machine learning techniques such as regression and classification to solve a variety of business problems using real-world data.
- Build strong intuitions about machine learning techniques by implementing them in a hands-on interactive programming environment.
- Systematically investigate and improve the results produced by machine learning models.
- Collaborate productively with your data science team.
- Keep up with the rapidly progressing field of machine learning and AI.



“The math is right. It’s just in poor taste.”

New Yorker, August 28, 2015

THE IMPORTANT QUESTIONS

- How do you take available data and translate it into a business problem you can solve?
- How do you know your machine learning system is not learning too little or too much?
- What's the right measure of a machine learning system's performance?

Understanding the business context is a critical component of success in building machine learning systems.

SCORING AND EVALUATION

- Homework assignments - 60%
- Group assignment/presentation – 20%
- Final exam – 20%

- Course materials will be available on Canvas/GitHub

TOUR OF *ORANGE*

https://orange.biolab.si/

Data Mining
Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

Download Orange