Deep Learning: A Critical Appraisal

Slides summarizing the paper by Gary Marcus

About this lecture

- This lecture is the first of a two-part series reviewing the paper ***Deep Learning: A Critical Appraisal*** by Gary Marcus

- We will review key points in the paper that align with course objectives

- We will also discuss how many of them apply to machine learning approaches outside of deep learning

Part 1

- Deep Learning is Data Hungry
- Limited capability for transfer
- Open-ended inference
- Lack of transparency
- Not integrated well with prior knowledge
- Cannot distinguish causation from correlation
- Deep learning assumes a stable world
- Difficult to engineer over the long-term

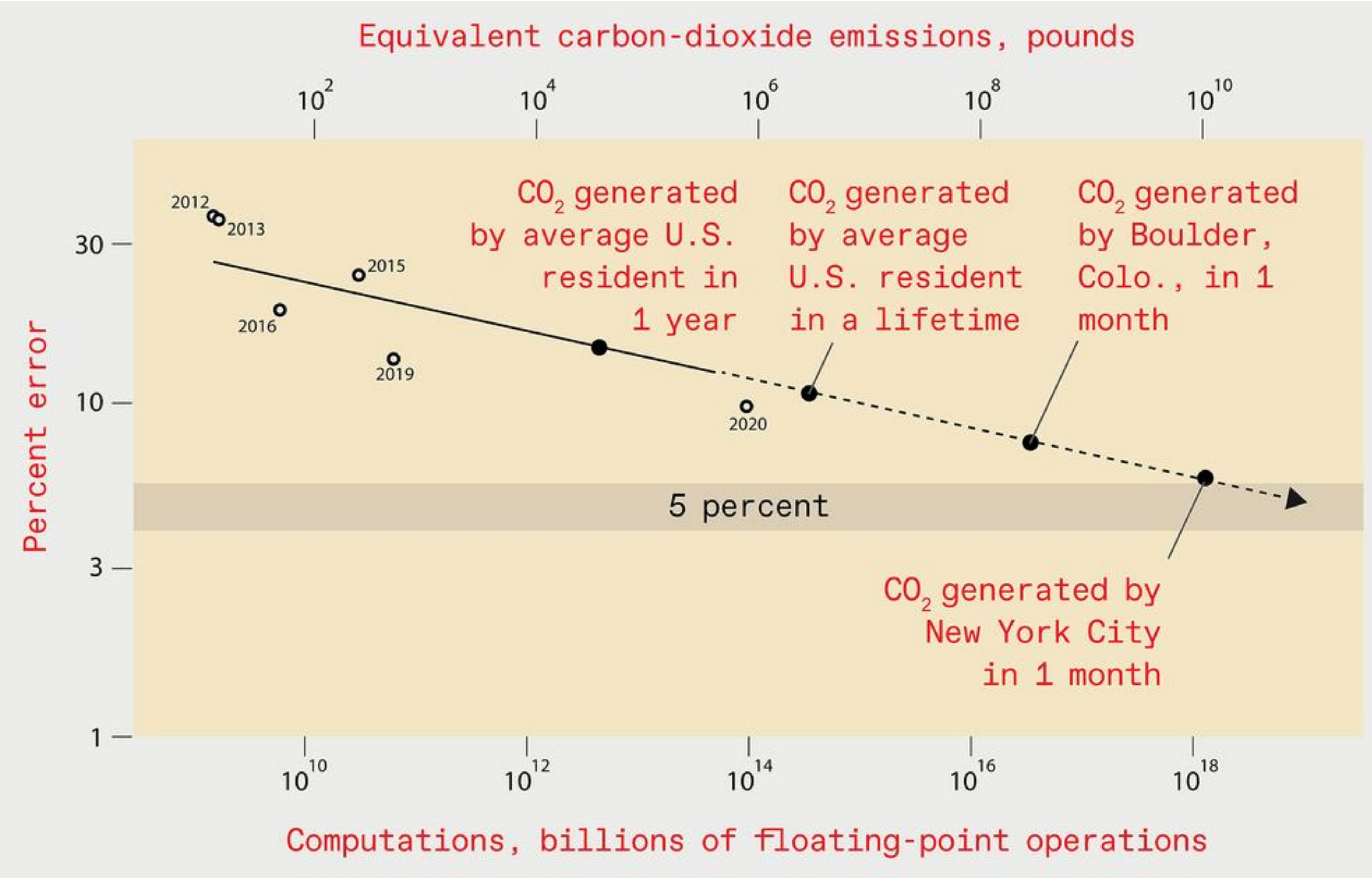
Deep Learning is Data Hungry

- Key deep learning intuition: More layers, more data provides better performance
  - Note: this was not true of earlier ML approaches
  - But comes at a cost

- A factor of *k* improvement would require k^2 samples and k^4 parameters

- Not just computation costs, but energy costs and carbon emissions

- This is because deep learning is inherently ***overparameterized*** - meaning the number of parameters exceeds the size of the training set



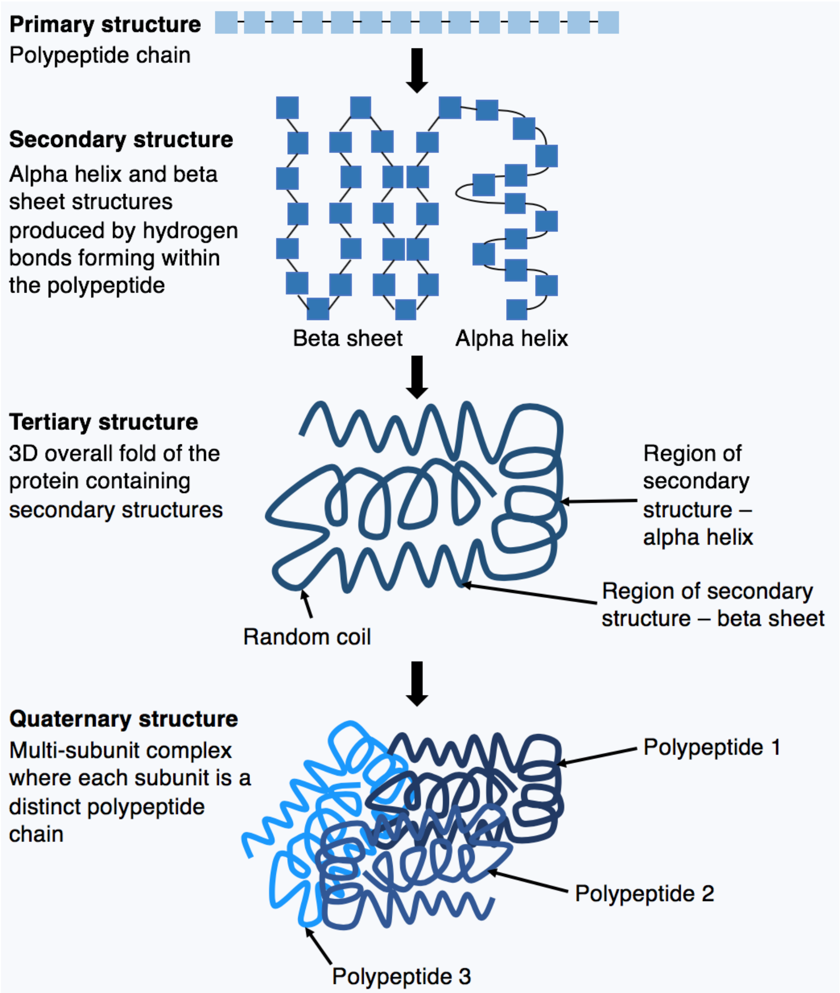
Source: Thompson et. al, IEEE Spectrum https://spectrum.ieee.org/deep-learning-computational-cost

Google/DeepMind AlphaFold/AlphaFold2: Parametric Approach "Solves" Protein Folding

- Predicting how a polypeptide chain folds is a challenging problem in biology
- DeepMind used a deep learning approach to solve this problem in a large variety of cases
- The model was trained on 29,000 proteins
- But uses 21 million parameters

More unknowns than samples – seems problematics

Ends up working well due to SGD's effectiveness



Note: These numbers refer to the original version of AlphaFold.

Deep Learning is Data Hungry

- DL’s data hunger contrasts with human ***data efficiency***
- Humans are able to generalize from small data, often just single examples
- We see this repeatedly with small children
- There are often cases where the data is just lacking (e.g. predicting the next pandemic) – which does not lend itself well to deep learning methods

Limited Capacity for Transfer

- Small perturbations can potentially cause wildly inaccurate results
- Counter examples shown for various systems that play video games, board games, and answered trivia questions that lowered accuracy
- Key intuition: the counter examples were not radically out of the distribution
- Why this happens: the deep learning models often learn superficial patterns which can be easily fooled

Limited Capacity for Transfer

- That said, certain networks (especially for image processing) exhibit transfer capability when added as part of the model and trained in a new domain (this is not what Gary Marcus is referring to when he discusses transfer)

- Note that this is a problem with other models, but is more pronounced with featureless approaches

- This problem also is related to Gary Marcus' 3rd and 9th points (inability to model hierarchies and untrustworthiness of results)

Open-ended inference

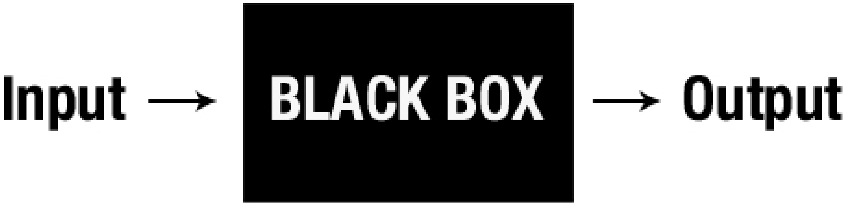
- Open-ended inferences (e.g. reading a text and answering arbitrary questions about characters intent) is not solved by deep learning

- Humans do well at this task, and without large amounts of training data

- Not solved by other means as well

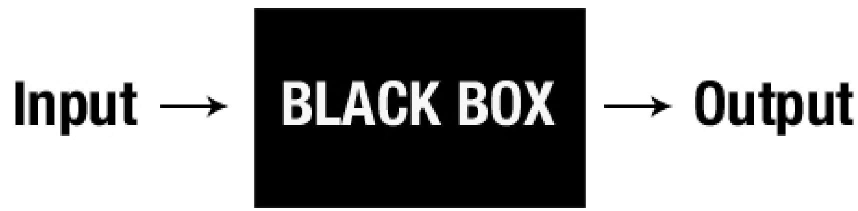
Explainability

- Deep learning networks are black box systems
  - The inner working are not understood by the user
- Large numbers of parameters and neurons prohibit the deciphering the steps a model takes to get to a result



Explainability

- This directly relates to trust in results, ability to determine bias in a system, and transfer
- Other machine learning methods are explainable
  - Rule mining
  - Decision tree learning



Part 2

- Deep Learning is Data Hungry
- Limited capability for transfer
- Open-ended inference
- Lack of transparency
- Not integrated well with prior knowledge
- Cannot distinguish causation from correlation
- Deep learning assumes a stable world
- Difficult to engineer over the long-term

Not Integrated with Prior Knowledge

Gary Marcus gives some examples of items humans make inferences without prior knowledge (but use of “common sense knowledge”)

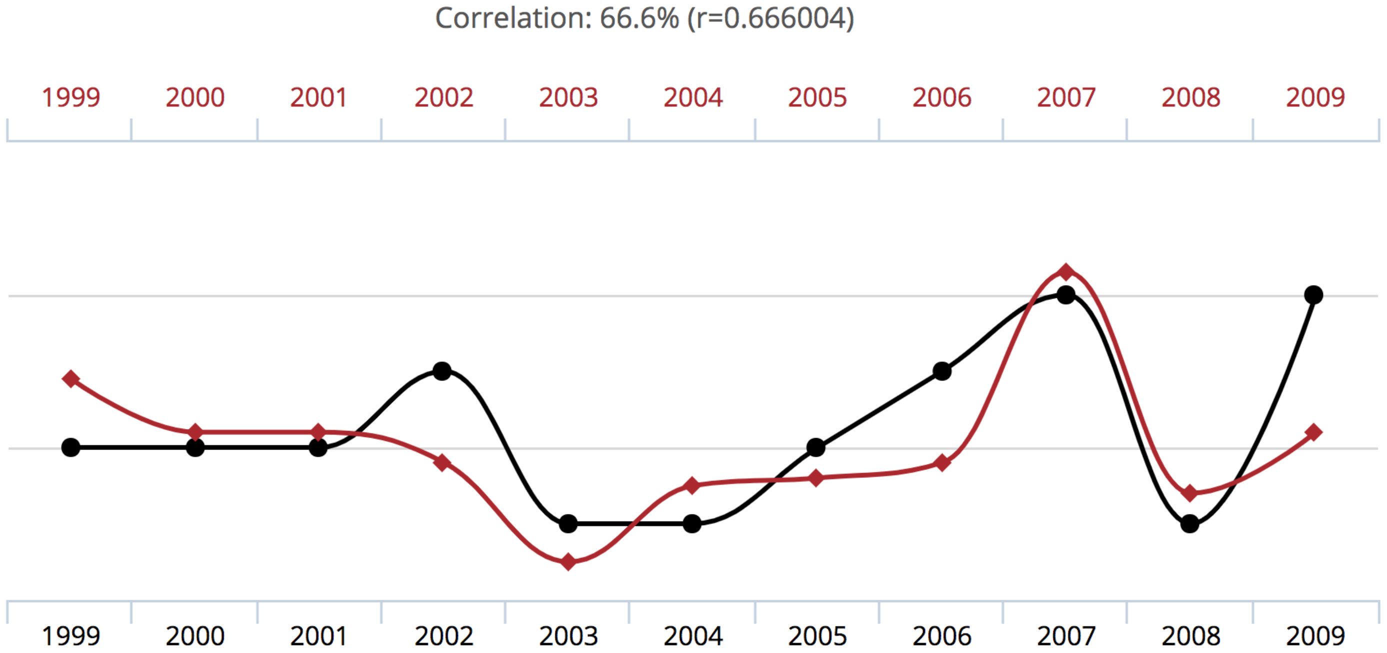
- Who is taller, Prince William or his baby son Prince George?
- Can you make a salad out of a polyester shirt?
- If you stick a pin into a carrot, does it make a hole in the carrot or in the pin?

Note that for these questions, training on historical data does not make sense, but having prior knowledge does.

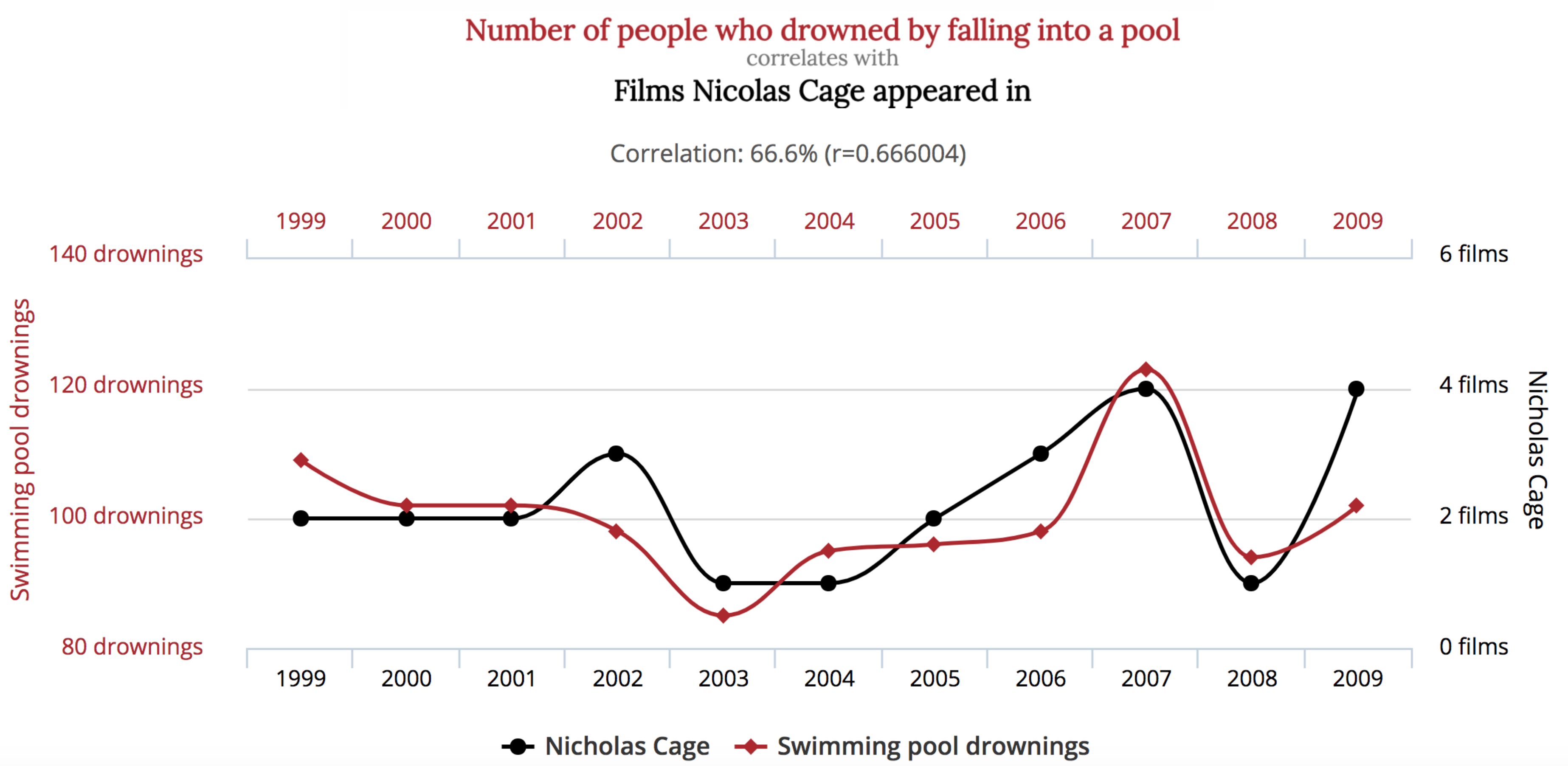
Not Integrated with Prior Knowledge

- Deep learning models are designed to learn from data, not pre-specified models
- Common sense knowledge and physical laws are two common examples of knowledge you may want to integrate into an ML system
- Since Marcus' paper, there have been some attempts to integrate existing knowledge, but this is still in the early phases
- Even so, the utility of such approaches is questionable when there is no explainable output

Correlation vs. Causation



Correlation vs. Causation

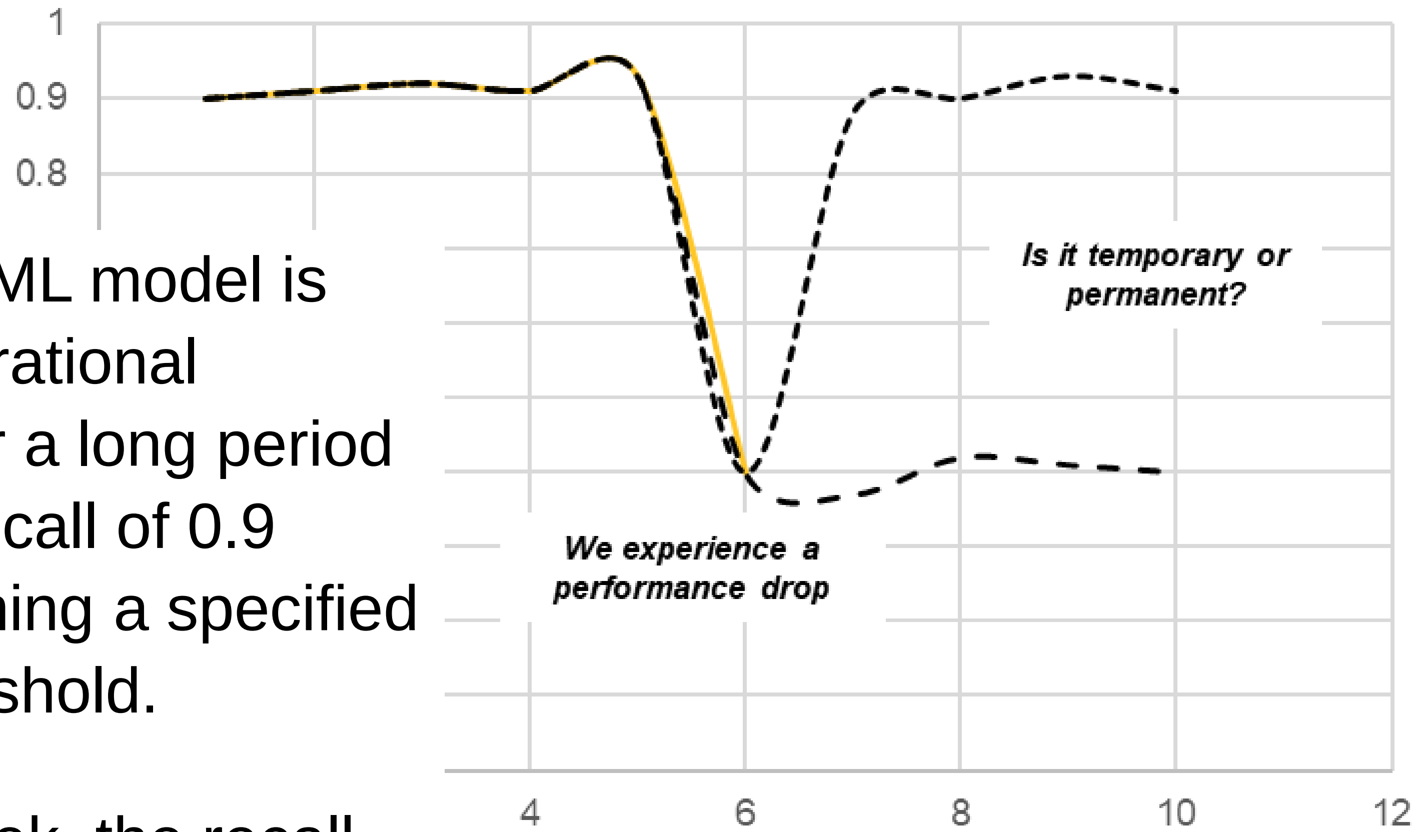


Source: http://www.tylervigen.com/spurious-correlations

Correlation vs. Causation

- The use of correlation is a problem across most popular machine learning algorithms
- It can become more pronounced in deep learning due to overparameterization
- Note that regularization does not address causality (e.g. the example of the last slide only depends on a single feature)

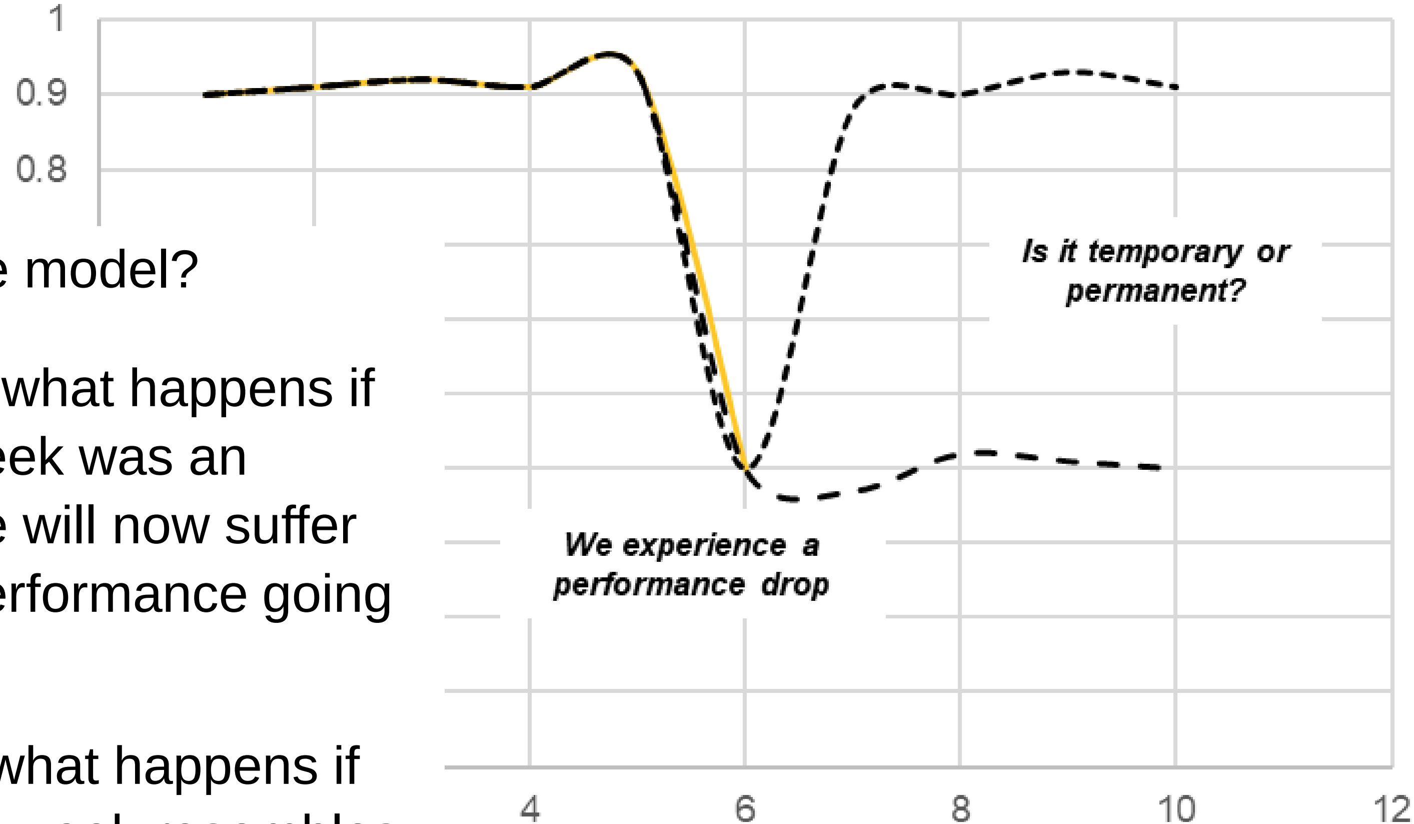
Assumption of a Stable World



Consider: an ML model is providing operational predictions for a long period of time with recall of 0.9 while maintaining a specified precision threshold.

For a new week, the recall drops to 0.4 – missing 3 of 5 predictions.

Assumption of a Stable World



Do we retrain the model?

If **<u>yes</u>** then what happens if the past week was an outlier – we will now suffer a loss of performance going forward

If **<u>no</u>** then what happens if the current week resembles a more permanent change in distribution – we will now suffer a loss of performance going forward

Assumption of a Stable World

- Deep learning is tied to the training data
- Changes in data distribution
- Issue with many machine learning algorithms, but more pronounced with deep learning due to large volumes of training data

Engineering Difficutlies

- Deep learning models are easy to build
- However, they are difficult to maintain over time
- Deep learning systems are not modular in the sense that guarantees for individual components can be used to provide guarantees for the system as a whole

Ira A. Fulton Schools of Engineering

Arizona State University