Machine Learning: Closing

Hao Tang

March 27, 2026

Modeling

- We have used the word “model” a lot in this course without defining it.
- What is a model?

A **model** is an imaginary artifact that describes the inner working of something.

A **model** is an imaginary artifact that describes the inner working of something.

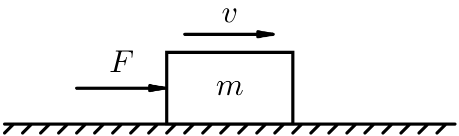
- imaginary

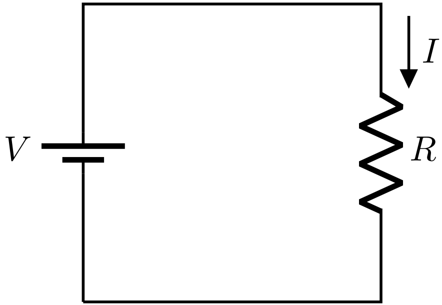
A **model** is an imaginary artifact that describes the inner working of something.

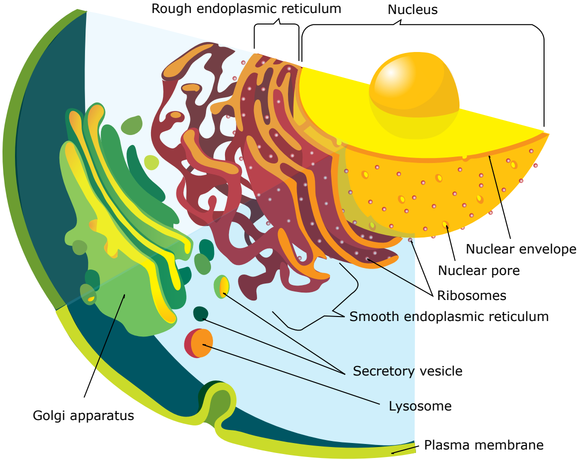
- imaginary
- describes inner working

A **model** is an imaginary artifact that describes the inner working of something.

- imaginary
- describes inner working
- of something


v
F
m


I
V
R


Rough endoplasmic reticulum
Nucleus
Nuclear envelope
Nuclear pore
Ribosomes
Smooth endoplasmic reticulum
Secretory vesicle
Lysosome
Golgi apparatus
Plasma membrane

The hypothesis class $\mathcal{H}$ is PAC learnable with an algorithm A if for any distribution $\mathcal{D}$, $\epsilon > 0$, and $0 < \delta < 1$, there exists N such that

$$\mathbb{P}_{S \sim \mathcal{D}^n}\left[L_{\mathcal{D}}(A(S)) - \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) > \epsilon\right] < \delta \tag{1}$$

for any $n \geq N$.

What about ...

- Perceptrons

What about ...

- Perceptrons
- Linear classifiers

What about ...

- Perceptrons
- Linear classifiers
- Gaussian mixture models

What about ...

- Perceptrons
- Linear classifiers
- Gaussian mixture models
- Transformers

What about ...

- Perceptrons
- Linear classifiers
- Gaussian mixture models
- Transformers
- Large language models

What is “artificial intelligence”?

What is “artificial intelligence”?

- Building models of intelligence, not solving tasks
- Involving philosophy, psychology, linguistics, etc

- Machine learning does not eliminate the need to build models.

- Machine learning does not eliminate the need to build models.
- What is machine learning?

- Machine learning does not eliminate the need to build models.
- What is machine learning?
- When to use machine learning?

- Machine learning does not eliminate the need to build models.
- What is machine learning?
- When to use machine learning?
- How to do use machine learning?

Types of thinking

- Mathematical
  - What happens if we have the optimal solution?
- Computational
  - What are the exact steps?
- Statistical
  - How many samples do we need?

Connections to other courses

- Foundations of Data Science (FDS)
- Applied Machine Learning (AML)
- ~~Machine Learning and Pattern Recognition (MLPR)~~
- ~~Probabilistic Modeling and Reasoning (PMR)~~
  $\rightarrow$ Probabilistic Machine Learning
- Machine Learning Practical (MLP)
- ~~Machine Learning Theory (MLT)~~
  $\rightarrow$ Advanced Topics in Machine Learning (ATML)
- ~~Reinforcement learning (RL)~~
  $\rightarrow$ Robot and Reinforcement Learning (RRL)

- Foundations of Natural Language Processing (FNLP)
- Accelerated Natural Language Processing (ANLP)
- ~~Natural Language Understanding, Generation, and Machine Translation (NLU+)~~
  $\rightarrow$ Advanced Topics in Natural Language Processing (ATNLP)
- Speech Processing (in PPLS)
- Automatic Speech Recognition (ASR)
- Speech Synthesis (in PPLS)
- Computer Vision (CV)
- Machine Learning Systems (MLS)

Things we haven't talked about

Tasks we haven't talked about

- Information retrieval
- 3D reconstruction
- Text generation
- Protein folding

Evaluations we haven't talked about

- Mean average precision
- Receiver operating characteristic (ROC) curve
- Word error rates (WER)
- Bilingual evaluation understudy (BLEU) score
- Mean opinion score (MOS)

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.
- k nearest neighbor
  - A point is like its neighbors.
- Decision trees
  - Decisions are based on conditional statements.
- More dimensionality reduction
  - Points in space have structures.
- Bayesian approaches
  - Beliefs are updated and marginalized.

Learning settings we haven't talked about

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?
- Semi-supervised learning
  - What do we do with unlabeled data?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?
- Semi-supervised learning
  - What do we do with unlabeled data?
- Federated learning
  - Can learning be distributed?

Aspects of machine learning we haven't talked about

- Privacy
- Fairness
- Interpretability

A learning algorithm A is differentially private if for all data sets S and S^i that differs in the i-th sample,

$$\mathbb{P}[A(S) \in \mathcal{G}] < e^{\epsilon}\mathbb{P}[A(S^i) \in \mathcal{G}] + \delta \tag{2}$$

for any subset $\mathcal{G} \subseteq \mathcal{H}$ of hypothesis.

Various types of fairness

- An outcome Y should be independent of groups G. ($Y \perp G$)
- A score R should be independent of the groups G given the outcome Y. ($R \perp G | Y$)
- An outcome Y should be independent of the groups G given the score R. ($Y \perp G | R$)

Debugging machine learning systems

Debugging machine learning systems

- “I am getting a 40% accuray while this paper is getting 80%.”

Debugging machine learning systems

- "I am getting a 40% accuray while this paper is getting 80%."
- "But the loss is not coming down during training."

Debugging machine learning systems

- "I am getting a 40% accuray while this paper is getting 80%."
- "But the loss is not coming down during training."
- "How do I know if there is a bug in my code?"

Debugging machine learning systems

- “I am getting a 40% accuray while this paper is getting 80%.”
- “But the loss is not coming down during training.”
- “How do I know if there is a bug in my code?”
- “I have seen papers reporting better results after using XYZ.”

What we have learned in this course

What we have learned in this course

- Formalize a ML problem with math
- Read and understand ML theorems
- Turn ML algorithms into programs