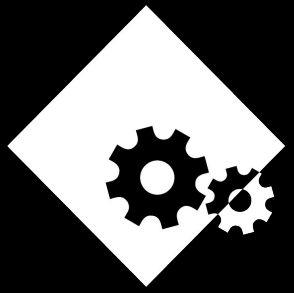


acm research.

build night 1

intros & onboarding

acm research.

welcome to acm research!

be proud. seriously.

this team is extremely stacked*:

Roman H. is very passionate about internet privacy and builds a lot of cool math and programming projects

Max H. is a prolific open source contributor, w/ over 764k lines on GitHub

Anh N. works with Dr. Gupta on AI research and is a recipient of the highly competitive GHC scholarship

Megan V. works on cybersecurity projects and is a recipient of the highly competitive GHC scholarship (she also works at the CSMC!!)

* this stuff is based on what I could find and is not exhaustive 😮

acm research.

agenda

- administrivia (team contract, slack/github)
- go over the structure of this project
- define what ML is
  define the next-word prediction problem
- discuss how homework works
- assign homework

 onboarding

project structure*

1. welcome & problem definition
2. intro to machine learning
3. sequence models
4. sequence models ii federated learning
5. federated learning ii
6. data sourcing & preparation
7. data prep ii & model training
8. model pruning & federation
9. poster & presentation work
10. poster work ii & practice

* this will probably change. hopefully not though because it makes planning things a pain. so yeah.

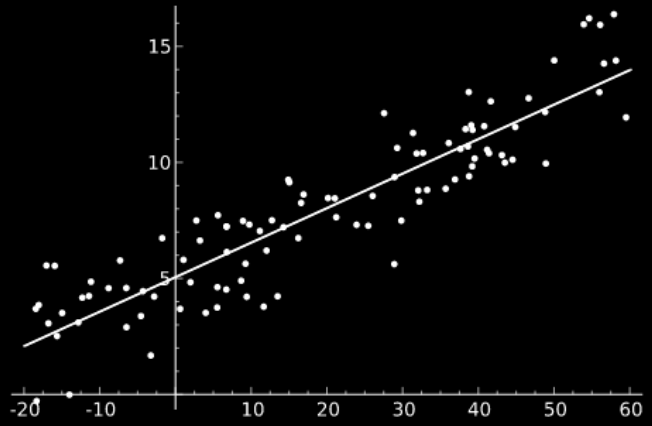
acm research.

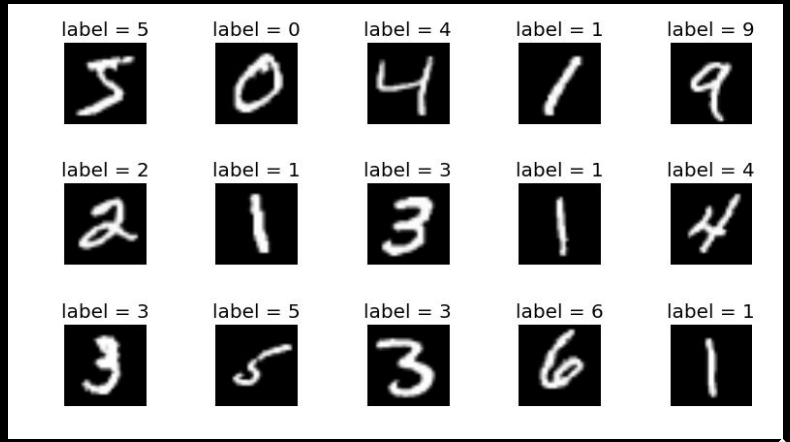
Defining Machine Learning

- An **algorithm** is a well-defined procedure to solve a well-defined problem
- A machine learning algorithm uses data to *learn* from experience
- What does **learning** mean?
  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.

Classes of ML Tasks *T*

- Classification: Learn a function $f : \mathbb{R}^n \rightarrow \{1, 2, ..., k\}$
  (learn *f* mapping an n-dimensional real vector to k classes)
- Regression: Learn a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$
  (learn *f* mapping an n-dimensional real vector to a real number)





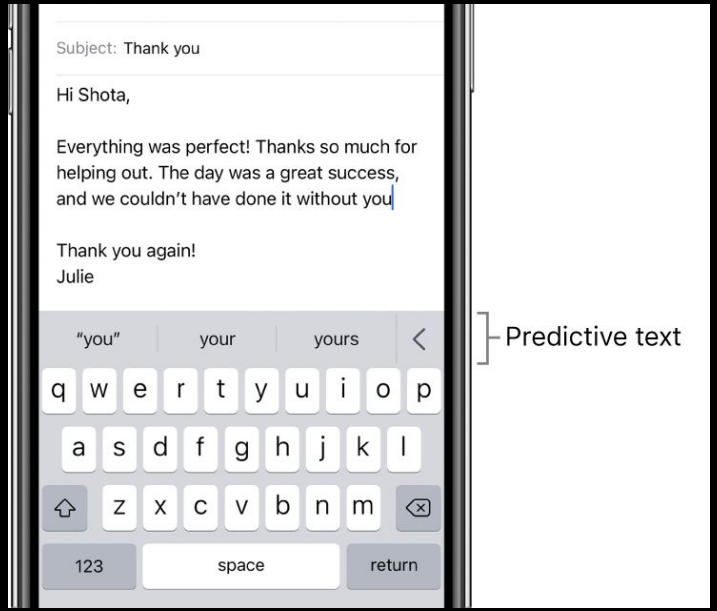
acm research.

Classes of ML Tasks *T*

- Clustering: Split *m* points in $\mathbb{R}^n$ into *k* groups
- Sequence modeling: Given the last *m* things appearing, what is the most probable next thing in the sequence?
  - The reading calls this the density estimation problem (this is a specialized case of it)

The Next-Word Prediction Problem & Our Project

- Given the last *m* words as context, predict the most probable next word in the sequence
- This model is fairly straightforward to build and there's a ton of literature on it (the generalized name for it is a Language Model)



The Next-Word Prediction Problem & Our Project

- We'll be training the model on our own, and then exploring ways to fine-tune it to language patterns using **federated learning**
- Federation allows for privacy to be preserved by training small portions of the model on-device based on user input
- There are many different ways to then take those small updates and aggregate them

how homework works

- homework will consist of working on your project, a coding exercise, or readings. all homework will be posted on github in the notes.
- doing these are crucial to conceptual understanding and your successful completion of the project!
- this is a minimum set of readings. feel free to explore the topic using resources you find online.
- these first few weeks will be reading and exercise heavy! once you start working on your model, readings/exercises will cool off
- this emulates how in a lab you may be reading papers or working on small experiments to understand concepts

👀 homework?