

A Tiny Large Model

CSC2541H1 Topics in Machine Learning, Winter 2025, UToronto

Chris J. Maddison

Announcements

- Comment on last class
- Presentations
 - If you were assigned to Jan 31, please come see me after class
- Change in office hours time
 - 1-2PM Fridays
- We are recording lectures and you can find the recordings on Quercus
 - See OCCS Student App tab

Recap

- We covered some key concepts last week.
 - **Prediction** – specifying the probability of an event and scoring your performance
 - **Learning** – procedures that estimate predictors from data
 - **Conditional prediction** – computing predictions and learning with additional data called “features”
 - **Large models** – scaled up non-linear, parametric predictors of internet text
- We focused on binary prediction and covered cross-entropy, empirical risk minimization, and logistic regression.

This week

- Implement these ideas!
- We will build a Transformer from scratch.
- The concepts that we covered last week form the foundation.
- Today's lecture is in JAX code.
 - We're assuming that you've at least seen numpy, Python's numerical computing library.
 - Ideally you've seen JAX or PyTorch.
 - If you haven't, please come to office hours to discuss! Happy to help.