INFO251 – Applied Machine Learning

Lab 3

Emily Aiken

Announcements

- **Problem Set 1** grades released
- **Problem Set 2** solutions posted
- **Problem Set 3** released, due Monday, February 21
- **Add/drop deadline** on Friday, February 18
- **GSI office hours**: Hybrid, South Hall Room 6a, **3:30 - 4:30**
  - I'll be answering questions in Physics Hall 3 after lab **3:00 - 3:30**

Today: Vectorized Computation

- Creating and manipulating matrices in Python
- Matrix operations: Addition, multiplication, dot product
- Efficient vectorized computation

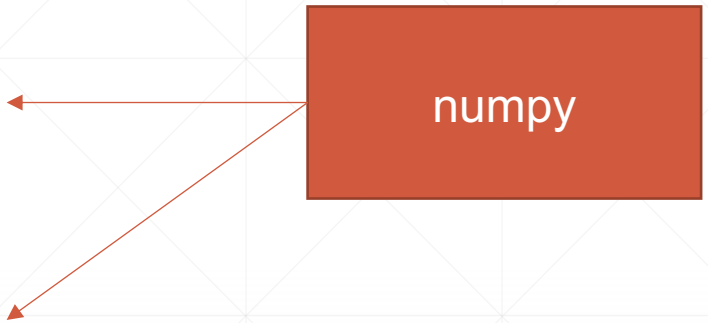
Today's programming tool: `numpy`

How to make a program run fast

- Programming language
  - **Fast:** C, C++, Java, Lisp/OCaml
  - **Slow:** Python
  - **Very slow:** R
- Writing efficient code
  - For loops vs. vectorized computation
- Hardware and parallelization
  - Run parts of a program in parallel on separate cores -- on a single machine or in a distributed system
  - Software packages for parallelizing data analysis in python: pyspark, dask
  - For more: **CS267**

How to make a program run fast

- Programming language
  - **Fast:** C, C++, Java, Lisp/OCaml
  - **Slow:** Python
  - **Very slow:** R
- Writing efficient code
  - For loops vs. vectorized computation



- Hardware and parallelization
  - Run parts of a program in parallel on separate cores -- on a single machine or in a distributed system
  - Software packages for parallelizing data analysis in python: pyspark, dask
  - For more: **CS267**