Welcome to CSE 427
Computational Biology

Professor: Su-In Lee
CSE & Genome Sciences
Office hour: Mon 2-3:20 pm
Office: Paul Allen Center (CSE building) 536

Teaching assistants
Scott Lundberg, CSE PhD student (first 5 weeks)
Safiye Celik, CSE PhD student (next 5 weeks)
Office hour: Thu 10-11:45am @ PAC 220

CS and biology seem to have very different goals and methodologies

Introducing teaching staff
Biology is an information science

- DNA acts as the “brain” of the cell, telling the cell how to properly grow and work.

Each individual has a slightly different version of a DNA sequence

- Each species has a unique DNA sequence.

DNA tells the cell how to perform various tasks in a cell

- Gene regulation
- Gene expression
- Gene interaction map ("social network" of genes)

Gene (>20,000 in human)
Biological information in the 21st century

- DNA sequence: >1M letters known to differ among individuals.
- RNA expression levels of 20K genes
- Protein levels of 20K genes
- Each DNA letter can be biochemically modified
  - How many numbers? Definitely >1M!

Modern biology is about mining very large, complex data

- Interesting questions arise such as
  - How the DNA sequences are different between human and chimpanzee?
  - Which parts in the DNA sequence determine susceptibility to dementia?
  - How epigenetic modifications to DNA influence cell functions?
  - How the “social network” of genes are different between cancer and normal cells?

CS and Statistics play a key role!

Course structure

- Topics to be discussed
  - Part 1: Sequence analysis
  - Part 2: Genetics
  - Part 3: Epigenomics
  - Part 4: Systems biology

- Methods broadly applicable to other applications
  - Dynamic programming, expectation maximization, hidden Markov models, Bayesian networks, clustering, classification
Prerequisites

- No background in biology is required
- Familiar with:
  - Basic statistics
  - Probability theory
  - Basic ML methods (e.g., linear regression) (optional)
- Programming skills in MATLAB, R, JAVA, C++, Perl, or Python
- Please contact us if you are unsure of our expectations

Logistics

- Course website: [http://cs.washington.edu/427](http://cs.washington.edu/427)
- Lecture notes, readings, HW assignments, project
- Due dates for homework and project
- Discussion board, HW dropbox and gradebook
- Mailing list: [cse427a_wi15@uw.edu](mailto:cse427a_wi15@uw.edu)

Course resources

- Lecture notes will become available in the morning before the class

Lecture notes

- Lecture slides will become available in the morning before the class
Readings

- No required textbook
- Readings will include papers and books listed on course website
- Biological background
  - The Cell, a molecular approach by Copper
  - Genetics, from genes to genomes by Hartwell and more
  - Principles of Population genetics by Hartl & Clark
- Computational background
  - Probabilistic graphical models by Daphne Koller & Nir Friedman
  - Andrew Ng's machine learning lecture note (cs229.stanford.edu)

Homework

- 4 homework assignments
  - HW will involve programming and written tasks.
  - One project-like homework (HW #4)
  - Will become available by 11:59pm on the assigned date
- Due at 11:59pm on the assigned due date
- Allowed 3 free late (calendar) days
- Collaboration allowed, but write-ups and coding must be done individually
  - Please write down team members’ names

Grading

- Homework assignments (90%)
  - HW #1, #2 and #3 (20% each)
  - HW #4 (30%)
- Attendance/participation (10%)
  - Good questions/answers in class

Course schedule