

---

# CSE 546, Autumn 2018 Machine Learning

Lecture: Tuesday, Thursday 11:30-12:50 Room: KNE 220 (<https://www.washington.edu/classroom/KNE+220>)

Instructor: Professor Kevin Jamieson (<http://homes.cs.washington.edu/~jamieson/>)

Contact: [cse546-instructors@cs.washington.edu](mailto:cse546-instructors@cs.washington.edu)

Discussion: We will be using Mattermost, a secure Slack clone (invite link ([https://mattermost.cs.washington.edu/signup\\_user\\_complete/?id=41rkawd1apnztDXmjzyu6bzsno](https://mattermost.cs.washington.edu/signup_user_complete/?id=41rkawd1apnztDXmjzyu6bzsno)) works if you're registered, email instructors for access otherwise)

Office Hours (check discussion board for exceptions):

- TA, Jifan Zhang ([jifan@uw](mailto:jifan@uw)), Monday 3:30-4:30 PM, CSE 4th floor breakout
- TA, An-Tsu Chen ([atc22@uw](mailto:atc22@uw)), Wednesday 4:00-5:00 PM, CSE 220
- TA, Pascal Sturmfels ([psturm@uw](mailto:psturm@uw)), Wednesday 9:00AM-10:00 AM, CSE 007
- TA, Beibin Li ([beibin@uw](mailto:beibin@uw)), Wednesday 1:30-2:30 PM, CSE 220
- TA, Alon Milchgrub ([alonmil@uw](mailto:alonmil@uw)), Thursday 10:00-11:00AM, CSE 220
- TA, Kung-Hung (Henry) Lu ([henrylu@uw](mailto:henrylu@uw)), Friday 12:30-1:30 PM, CSE 007
- Instructor, Tuesday 4:00-5:00 PM, CSE 666

---

## About the Course and Prerequisites

Machine learning explores the study and construction of algorithms that can learn from historical data and make inferences about future outcomes. This study is a marriage of algorithms, computation, and statistics so this class will have healthy doses of each. The goals of this course are to provide a thorough grounding in the fundamental methodologies and algorithms of machine learning.

Prerequisites: Students entering the class should be comfortable with programming and should have a pre-existing working knowledge of linear algebra (MATH 308), vector calculus (MATH 324), probability and statistics (MATH 394/STAT390), and algorithms. For a brief refresher I recommend you consult the linear algebra and statistics/probability reference materials below.

---

## Textbook and reference materials

I will assign reading out of the following texts because they are excellent and their PDFs are offered for **free** by the authors.

- [HTF] The Elements of Statistical Learning: Data Mining, Inference, and Prediction (<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>), Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- [EH] Computer Age Statistical Inference: Algorithms, Evidence and Data Science ([https://web.stanford.edu/~hastie/CASI\\_files/PDF/casi.pdf](https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf)), Bradley Efron, Trevor Hastie.

If you buy one ML book, I would recommend HTF of above. If you buy an additional ML book, I would recommend Shalev-Schwartz and Ben-David of below.

You may also find these reference materials useful throughout the quarter.

- Machine Learning
  - Understanding Machine Learning: From Theory to Algorithms (<https://www.amazon.com/Understanding-Machine-Learning-Theory-Algorithms/dp/1107057132>), Shai Shalev-Shwartz, Shai Ben-David. A gentle introduction to theoretical machine learning.
  - Machine Learning: A Probabilistic Perspective (<https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020>), Kevin Murphy. A more Bayesian approach to machine learning.
- Linear Algebra and Matrix Analysis
  - Linear Algebra Review and Reference (<http://www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf>) by Zico Kolter and Chuong Do (free). Light refresher for linear algebra and matrix calculus if you're a bit rusty.
  - Linear Algebra (<https://www.math.ucdavis.edu/~linear/linear-guest.pdf>), David Cherney, Tom Denton, Rohit Thomas and Andrew Waldron (free). Introductory linear algebra text.

- Matrix Analysis (<http://a.co/d/2zrXXHb>) Horn and Johnson. A great reference from elementary to advanced material.
- Probability and Statistics
  - All of Statistics (<https://link.springer.com/book/10.1007/978-0-387-21736-9>), Larry Wasserman. Chapters 1-5 are a great probability refresher and the book is a good reference for statistics.
  - A First Course in Probability (<http://www.mypearsonstore.com/bookstore/first-course-in-probability-9780134753119?xid=PS2D>), Sheldon Ross. Elementary concepts (previous editions are a couple bucks on Amazon)
- Optimization
  - Numerical Optimization (<https://link.springer.com/book/10.1007%2F978-0-387-40065-5>), Nocedal, Wright (SpringLink: free on UW network). Practical algorithms and advice for general optimization problems.
  - Convex Optimization: Algorithms and Complexity (<http://sbubeck.com/Bubeck15.pdf>), Sébastien Bubeck. Elegant proofs for the most popular optimization procedures used in machine learning.
- Python
  - [www.learnpython.org](https://www.learnpython.org/) (<https://www.learnpython.org/>) "Whether you are an experienced programmer or not, this website is intended for everyone who wishes to learn the Python programming language."
  - NumPy for Matlab users (<https://docs.scipy.org/doc/numpy/user/numpy-for-matlab-users.html>)
- Latex
  - Learn Latex in 30 minutes ([https://www.sharelatex.com/learn/Learn\\_LaTeX\\_in\\_30\\_minutes](https://www.sharelatex.com/learn/Learn_LaTeX_in_30_minutes))
  - Overleaf (<https://www.overleaf.com/>). An online Latex editor.
  - Standalone Latex editor (<https://www.latex-project.org/get/>) on your local machine
  - Latex Math symbols (<http://web.ift.uib.no/Teori/KURS/WRK/TeX/symALL.html>)
  - Detexify (<http://detexify.kirelabs.org/classify.html>) LaTeX handwritten symbol recognition

## Discussion Forum and Email Communication

IMPORTANT: This class uses Mattermost (a secure Slack clone). An invite link will be available on the Canvas Discussion board. If not registered for the course, please request an invite link by sending an email to [cse546-instructors@cs.washington.edu](mailto:cse546-instructors@cs.washington.edu). All class announcements will be broadcasted on mattermost and you are responsible for keeping up to date on it (I suggest you turn on push notifications). The same applies to questions about homeworks, projects and lectures. Mattermost lowers the barrier to asking for help and encourages more interaction. It is also a place where students who are not registered can interact with the rest of the class (unlike Canvas). Please ask all course-related questions in a public channel on Mattermost as other students will often have the same question, or know the answer. If you have a question of personal matters, please email the instructors list: [cse546-instructors@cs.washington.edu](mailto:cse546-instructors@cs.washington.edu).

## Grading and Evaluation

Your grade will be based on 5 homework assignments (65%) and a final project (35%).

### Homework

Your homework score will be the smaller of 100 points and the cumulative number of points you receive on the assignments. The first homework is worth 10 points, and the final four are worth 25 each. This means if you receive grades  $(x_0, x_1, x_2, x_3, x_4)$  you will receive a score of  $\min$ . In particular, if you receive grades

- (10,25,25,25,0) you will get a total homework score of 85.
- (10,25,25,25,15) you will get a total homework score of 100.
- (10,25,25,25,25) you will get a total homework score of 100.

Homeworks must be submitted by the posted due date at **11:59 PM Seattle time**.

- **Late work will receive a score of 0.**
- **All assignments must be submitted (even if late for a score of 0). If not, you will not pass.**
- **All assignments are to be submitted electronically on canvas.**

Each homework assignment contains both theoretical questions and will have programming components.

- **You are required to use Python for the programming portions.** There are a number of Python resources above. You may use any numerical linear algebra package (e.g., NumPy/SciPy), but **you may not use machine learning libraries (e.g. sklearn, pytorch, tensorflow)** unless otherwise specified (later in the course). Your analysis and code should all be included in a single PDF, with your code at the end very end.

- **You must submit your HW as a typed PDF document typeset in Latex (not handwritten).** There are a number of Latex resources above. Also note that LaTeX is installed on department-run machines.

The first homework (10 points) is designed to be a **review** of the course prerequisites. If this assignment requires significant effort (e.g., several hours) or contains unfamiliar topics, you should strongly consider dropping the course and revisiting the prerequisites. Its secondary purpose is to get you comfortable with Python and Latex.

**COLLABORATION POLICY:** Homework must be done individually: each student must submit their own answers. In addition, each student must write and submit their own code in the programming part of the assignment (we may run your code). It is acceptable, however, for students to collaborate in figuring out answers and helping each other solve the problems. You must also indicate on each homework with whom you collaborated.

**RE-GRADING POLICY:** All requests for regrading should be submitted to Gradescope directly. Office hours and in person discussions are limited solely to asking knowledge related questions, not grade related questions. If you feel that we have made an error in grading your homework, please let us know with a written explanation, and we will consider the request. Please note that regrading of a homework means the entire assignment may be regraded which may cause your grade on the entire homework set to go up or down.

**LATE POLICY:** Homeworks must be submitted online by the posted due date. With the exception of the poster presentation, all work is to be submitted online. **There is no credit for late work.** The homework scoring system of above is an attempt to minimize the rigidity of this policy. We may make special arrangements for alternative dates for poster presentation (contact the instructors). If you are unable to meet the deadlines due to travel, conferences, other deadlines, or any other reason, do not enroll in the class.

**HONOR CODE:** As we sometimes reuse problem set questions from previous years, covered by papers and webpages, we expect the students not to copy, refer to, or look at the solutions in preparing their answers (referring to unauthorized material is considered a violation of the honor code). Similarly, we expect students not to google directly for answers. The homework is to help you think about the material, and we expect you to make an honest effort to solve the problems. If you do happen to use other material, it must be acknowledged clearly with a citation on the submitted solution. For more information, please see the CSE Academic Misconduct policy (<https://www.cs.washington.edu/students/policies/misconduct>) that this course adheres to.

## Project

You will work independently or with a partner on a machine learning project spanning most of the quarter ending with a poster presentation and written report. You may use techniques developed in this course but are also encouraged to learn and apply new methods. The project should address a novel question with a non-obvious answer and must have a real-data component. We will provide some seed project ideas. You can pick one of these ideas, and explore the data and algorithms within and beyond what we suggest. You can also use your own data/ideas, but, in this case, you have to make sure you have the data available at the time of the proposal and a nice roadmap, since a quarter is too short to explore a brand new concept. The components of the project are

- **Project Proposal (10 points):** A one page maximum description of your project with: 1) project title, 2) dataset(s), 3) Project idea (two paragraphs), 4) Software you will write and/or use, 5) papers to read (include 1-3 relevant papers), 6) will you have a teammate?, and 7) what will you complete by the milestone (experimental results are expected)?
- **Project Milestone (15 points):** Your write up should be 3 pages maximum (not including references) in Camera-ready NIPS format (<https://nips.cc/Conferences/2018/PaperInformation/StyleFiles>). You should describe the results of your first experiments here and what you wish to accomplish before the final presentation and paper submission. Note that, as with any conference, the page limits are strict! Papers over the limit will not be considered.
- **Poster presentation (15 points):** We will hold a poster session in the Atrium of the Paul Allen Center. Each team will be given a stand to present a poster summarizing the project motivation, methodology, and results. The poster session will give you a chance to show off the hard work you put into your project, and to learn about the projects of your peers. We will provide poster boards that are 32x40 inches. Both one large poster or several pinned pages are OK (fonts should be easily readable from 5 feet away).
- **Project Report (60 points):** Your write up should be 4 pages maximum (not including references) in Camera-ready NIPS format (<https://nips.cc/Conferences/2018/PaperInformation/StyleFiles>). You may have unlimited appendices for clarifications, however, no reviewer is required to look at these to evaluate the work. You should describe the task you solved, your approach, the algorithms, the results, and the conclusions of your analysis. Note that, as with any conference, the page limits are strict! Papers over the limit will not be considered.

Example project ideas can be found here ([project.html](#)).

# Homework

- Homework 0: Warm up (10 points)
  - Due: 11:59 PM Thursday October 4
  - Homework: PDF (resources/hw0.pdf), LaTeX (resources/hw0.zip)
- Homework 1: MLE, Bias-variance, Ridge Regression (25 points)
  - Due: 11:59 PM Thursday October 18
  - Homework: PDF (resources/hw1.pdf), LaTeX (resources/hw1.tex)
- Homework 2: Empirical Risk Minimization, Lasso, Logistic regression (25 points)
  - Due: Thursday November 1
  - Homework: PDF (resources/hw2.pdf), LaTeX (resources/hw2.tex), data for problem 4 (resources/yelp\_data.zip)
- Homework 3: Bayesian inference, Kernel Regression, K-means, Matrix completion (25 points)
  - Due: Tuesday November 20
  - Homework: PDF (resources/hw3.pdf), LaTeX (resources/hw3.tex), data for problem 5 (removed)
- Homework 4: EM, Convex programming, Neural networks (25 points)
  - Due: Tuesday December 4
  - Homework: PDF (resources/hw4.pdf), LaTeX (resources/hw4.tex)
- Homework 3, problem 5 revisited **optional**: (see assignment)
  - Due: Tuesday December 12
  - Homework: PDF (resources/hw3\_optional.pdf), LaTeX (resources/hw3\_optional.tex), data for this problem (resources/hw3p5\_data.zip)

## Important Dates

| Date                | Deliverable Due                      |
|---------------------|--------------------------------------|
| 10/4                | Homework 0                           |
| 10/18               | Homework 1                           |
| 10/25               | Project proposal                     |
| 11/1                | Homework 2                           |
| 11/15               | Project milestone                    |
| 11/20               | Homework 3                           |
| 12/4                | Homework 4                           |
| 12/4, 4:30-7:30 PM  | Poster presentation                  |
| 12/7                | Project report due                   |
| 12/12, 4:30-7:30 PM | Poster presentation                  |
| 12/12               | <b>Optional</b> Homework 3 revisited |
| 12/14               | Project Reviews due                  |

## Schedule

- Lecture 1 (9/28)
  - Topics: Welcome/overview, MLE for Bernoulli and Gaussians
  - Reading: HTF 1, 3.1-3.2; EH 4-4.2
  - Additional reading: Wasserman 9.3-9.7
  - Slides: lecture slides (resources/lecture\_1.pdf), annotated lecture slides (resources/lecture\_1\_annotated.pdf)
- Optional linear algebra and probability review (10/1)

- Monday 5:00-7:00 PM, ARC 147
- Review PDF (resources/review.pdf)
- Lecture 2 (10/2)
  - Topics: Linear Least Squares, Bias-Variance tradeoff
  - Reading: HTF 2.5, 3.1-3.2, 7.1-7.3, 3.4
  - Slides: lecture slides (resources/lecture\_2.pdf), annotated lecture slides (resources/lecture\_2\_annotated.pdf)
- Lecture 3 (10/4)
  - Topics: Bias-Variance tradeoff, Ridge regression
  - Reading: HTF 7.1-7.3, 7.10-7.12, 3.4
  - Slides: lecture slides (resources/lecture\_3.pdf), annotated lecture slides (resources/lecture\_3\_annotated.pdf)
- Lecture 4 (10/9)
  - Topics: k-fold cross validation, Lasso
  - Reading: HTF 7.1-7.3, 7.10-7.12, 3.4, 3.8.5-3.8.6
  - Slides: lecture slides (resources/lecture\_4.pdf), annotated lecture slides (resources/lecture\_4\_annotated.pdf)
- Lecture 5 (10/11)
  - Topics: Lasso, Logistic Regression
  - Reading: HTF 3.4, 3.8.5-3.8.6, 4.1-4.2, 4.4
  - Slides: lecture slides (resources/lecture\_5.pdf), annotated lecture slides (resources/lecture\_5\_annotated.pdf)
- Lecture 6 (10/16)
  - Topics: Logistic Regression, Optimization basics
  - Reading: HTF 4.1-4.2, 4.4
  - Additional reading: Nocedal and Wright 2-3
  - Slides: lecture slides (resources/lecture\_6.pdf), annotated lecture slides (resources/lecture\_6\_annotated.pdf)
- Lecture 7 (10/18)
  - Topics: Optimization
  - Reading: HTF 4.4
  - Additional reading: Nocedal and Wright 2-3
  - Slides: lecture slides (resources/lecture\_7.pdf), annotated lecture slides (resources/lecture\_7\_annotated.pdf)
- Lecture 8 (10/23)
  - Topics: Perceptron, SVM, Bootstrap
  - Reading: HTF 4.5, 12-12.2; EH 10-10.4, 11-11.2
  - Slides: lecture slides (resources/lecture\_8.pdf), annotated lecture slides (resources/lecture\_8\_annotated.pdf)
- Lecture 9 (10/25)
  - Topics: Bootstrap, Generative/Discriminative, hypothesis testing
  - Reading: HTF 4.1-4.3.1, 18.7; EH 2-2.2, 10-10.4, 11-11.2
  - Slides: lecture slides (resources/lecture\_9.pdf), annotated lecture slides (resources/lecture\_9\_annotated.pdf)
- Lecture 10 (10/30)
  - Topics: hypothesis and multiple testing, Bayesian methods
  - Reading: HTF 18.7; EH 10-10.4, 11-11.2, 3
  - Additional reading: Wasserman 11
  - Slides: lecture slides (resources/lecture\_10.pdf), annotated lecture slides (resources/lecture\_10\_annotated.pdf)
- Lecture 11 (11/1)
  - Topics: Bayesian methods, Nearest Neighbors, Kernels
  - Reading: HTF 2.5, 2.8, 6.1-6.3; EH 3
  - Slides: lecture slides (resources/lecture\_11.pdf), annotated lecture slides (resources/lecture\_11\_annotated.pdf)
- Lecture 12 (11/6)
  - Topics: Kernels, PCA
  - Reading: HTF 5.8, 12.3; 14.5
  - Slides: lecture slides (resources/lecture\_12.pdf), annotated lecture slides (resources/lecture\_12\_annotated.pdf)
- Lecture 13 (11/8)
  - Topics: PCA, SVD
  - Reading: HTF 14.5, 14.3
  - Slides: lecture slides (resources/lecture\_13.pdf), annotated lecture slides (resources/lecture\_13\_annotated.pdf)
- Lecture 14 (11/13)

- Topics: Matrix completion, K-means
- Reading: HTF 14.5, 14.3
- Slides: lecture slides (resources/lecture\_14.pdf)
- Lecture 15 (11/15)
  - Topics: K-means, Mixture models, EM
  - Reading: HTF 14.5, 14.3, 8.5
  - Slides: lecture slides (resources/lecture\_15.pdf), annotated lecture slides (resources/lecture\_15\_annotated.pdf)
- Lecture 16 (11/20)
  - Topics: Text and Image featurization, Hyperparameter tuning
  - Reading
    - Image convolutional networks blog post (<https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>)
    - Coates and Ng (2012) random patches ([https://www-cs.stanford.edu/~acoates/papers/coatesng\\_nntot2012.pdf](https://www-cs.stanford.edu/~acoates/papers/coatesng_nntot2012.pdf))
  - Slides: lecture slides (resources/lecture\_16.pdf), annotated lecture slides (resources/lecture\_16\_annotated.pdf)
- Lecture 17 (11/27)
  - Topics: Image featurization, Hyperparameter tuning, Neural Networks
  - Reading
    - Deep Learning book (<http://www.deeplearningbook.org/>)
  - Slides: lecture slides (resources/lecture\_17.pdf), annotated lecture slides (resources/lecture\_17\_annotated.pdf)
- Lecture 18 (11/29)
  - Topics: Back Propagation, Random Forests, Boosting
  - Reading: HTF 9.2, 15-15.3, 10-10.11
    - Deep Learning book (<http://www.deeplearningbook.org/>)
    - Automatic Differentiation in Machine Learning: a Survey (<https://arxiv.org/pdf/1502.05767.pdf>)
  - Slides: lecture slides (resources/lecture\_18.pdf), annotated lecture slides (resources/lecture\_18\_annotated.pdf)
- Lecture 19 (12/4)
  - Topics: Fairness, Boosting, PAC learning
  - Reading: HTF 10-1.10, 7.8-7.9
    - [www.fatml.org](http://www.fatml.org) (<http://www.fatml.org/>)
    - Attacking discrimination with smarter machine learning (<http://research.google.com/bigpicture/attacking-discrimination-in-ml/>)
  - Slides: lecture slides (resources/lecture\_19.pdf), annotated lecture slides (resources/lecture\_19\_annotated.pdf)
- Lecture 20 (12/6)
  - Topics: PAC Learning, No-Free Lunch Theorem, VC Dimension
  - Reading: HTF 7.8-7.9
    - Percy Liang's stat learning theory notes (<https://web.stanford.edu/class/cs229t/notes.pdf>)
  - Slides: lecture slides (resources/lecture\_20.pdf)