XCS229i: Machine Learning

Syllabus and Course Information

Welcome

Welcome to XCS229i: Machine Learning! This professional course is based on graduate-level material from Stanford's on-campus course, CS229, adapted for a professional certificate format and split into two parts (XCS229i: Machine Learning and XCS229ii: Machine Learning Strategy and Intro to Reinforcement Learning) to make the content and workload more manageable for working professionals. In this course you will:

- Learn from Stanford classroom lecture videos that have been edited and segmented by topic for easier navigation, reference, and review.
- Complete problem sets enhanced with additional supports and scaffolding.
- Receive support from Stanford-affiliated Course Facilitators.
- Connect with a cohort of peers from diverse locations and professional backgrounds.

Course Launch

All lecture videos will be available on the first day of the course **(April 12th) at 12:00pm Pacific Time**. Course problem sets will be released as indicated in the calendar below, without exception. Maintaining this schedule enables Course Facilitators to be most effective in providing support and answering questions on subject matter throughout the course.

Getting Started

This course will use different tools to distribute content, manage problem sets, and deliver support. They are:

- **SCPD Learning Management System** – accessed via the mystanfordconnection site which you used to apply to and enroll in this course.
- **GitHub** – to distribute code and data for the problem sets.
- **Slack** – for additional course support and class discussions.

Joining Slack

In addition to direct small group support from Course Facilitators (more details and guidelines in Course Facilitators, Support, and Guidelines section below), the cohort will have a Slack workspace to ask additional questions and discuss course topics**.** An email invitation to the Slack workspace will be sent to your email address on file with SCPD on **April 9th**.

If you have previously joined an SCPD/Stanford Slack Workspace for a previous course in the AI Professional Program, Slack does not send a notification when our staff 're-invites' you to this workspace. Instead, you are automatically re-activated, and on **April 9th** should proceed directly to http://xcs229i-scpd.slack.com/ → I have a guest account → Log in using your credentials.

Joining GitHub

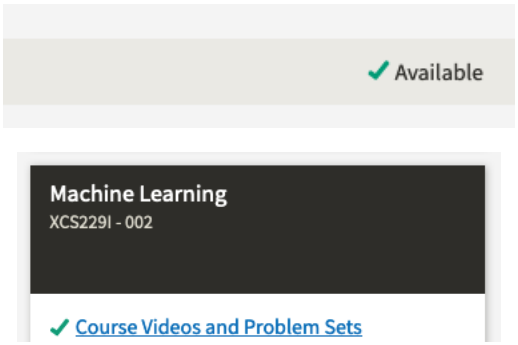
You will receive an email invitation to a GitHub team called "XCS229i-Spring-2021" at your address on file with SCPD. If you'd prefer that you receive a GitHub invitation at a different address, just let us know!

The team's repository will be blank to begin with – code files will be added gradually as each problem set is released (see schedule below).

Getting Started (On April 12)

Accessing Your Course

1. On **April 12th after 12pm Pacific Time**, log in to the mystanfordconnection account you used when applying for the XCS229i course.
2. XCS229i: Machine Learning will be visible as a live course. Click the link titled "Course Videos and Problem Sets" to enter our learning management system.



Course Calendar

Week	Suggested Videos	**Optional** Live Sessions (will also be recorded)	Problem Sets
Week 1 & 2 (Apr 12 - 25)	**Module 1:** *Linear Regression, Batch/Stochastic Gradient Descent, Normal Equation, Locally Weighted Regression, Probabilistic Interpretation of Linear Regression, Logistic Regression, Newton's Method*		PS 1 Released – Monday, Apr 12 PS 2 Released – Monday, Apr 12 PS 1 Due – Sunday, Apr 25
Week 3 & 4 (Apr 26 - May 9)	**Module 1:** *Perceptron, Exponential Family, GLM, Softmax Regression, GDA, Generative & Discriminative Analysis*	**Problem Set 1 Solution Walk-Through** Zoom - Times/Dates TBD	PS 3 Released – Friday, Apr 30 PS 2 Due – Sunday, May 9
Week 5 & 6 (May 10 - 23)	**Module 1:** *Naive Bayes, Laplace Smoothing & Event Models, SVM, Kernels, Decision Trees*	**Problem Set 2 Solution Walk-Through** Zoom - Times/Dates TBD	PS 4 Released – Friday, May 14 PS 3 Due – Sunday, May 23
Week 7 & 8 (May 24 - Jun 6)	**Module 1 & 2 & 3** *Ensemble Methods, Neural Networks, K-means Clustering, Mixture of Gaussians*	**Problem Set 3 Solution Walk-Through** Zoom - Times/Dates TBD	PS 5 Released – Friday, May 28 PS 4 Due – Sunday, June 6
Week 9 & 10 (Jun 7 - 20)	**Module 3** *Expectation-Maximization Algorithm, Factor Analysis, Independent Component Analysis*	**Problem Set 4 & 5 Solutions Walk-Through** Zoom - Times/Dates TBD	PS 5 Due – Sunday, June 20

Problem Sets and Grading

Problem sets will be released via the SCPD course platform on the dates noted above in the course calendar. Below is a brief summary of what each problem set will entail:

Problem Set 1

1. Convexity of Generalized Linear Models
2. Linear Regression: linear in what?

In this problem set, you will explore and derive the convexity property of Generalized Linear Models (GLMs). In the second problem, you will learn about different methods for hand-tuning linear regression.

Problem Set 2

1. A Linear Classifiers (logistic regression and GDA)
2. Poisson Regression

In this problem set, you will explore two probabilistic linear classifiers. First, a discriminative linear classifier: logistic regression. Second, a generative linear classifier: Gaussian discriminant analysis (GDA). Both the algorithms find a linear decision boundary that separates data into two classes, but each makes different assumptions. You will gain a deeper understanding of the similarities and differences (and, strengths and weaknesses) of these algorithms.

In the second problem, you will construct another kind of a commonly used GLM: Poisson Regression. You will start by showing that the Poisson distribution is in the exponential family, derive the functional form of the hypothesis, derive the update rules for training models, and finally train a model and make predictions.

Problem Set 3

1. Constructing Kernels

In this problem set, you will learn to construct kernels and gain a discerning eye for the general requirements for a function to be a valid kernel.

Problem Set 4

1. Spam classification
2. Neural Networks: MNIST image classification

In this problem set, you will use the naive-Bayes algorithm and an SVM to build a spam classifier. In the second problem, you will create a neural network to solve the classic data science problem of classifying the MNIST dataset.

Problem Set 5

1. Semi-supervised EM
2. K-Means for compression

In this problem, you will apply the K-means algorithm to lossy image compression, by reducing the number of colors used in an image. In the second problem, you will explore one of the ways in which the Expectation-Maximization algorithm can be adapted to the semi-supervised setting, where there are some labeled examples along with unlabeled examples.

Honor Code

Students will be asked to review and maintain the standards set forth by the Stanford Honor Code when completing problem sets in this course. You can review the section labeled *Violations of the Honor Code* for representative examples relevant to this course.

Grading

Coding Questions are graded automatically upon upload and will show your score. You can continue to re-submit up until the due date. To view an example of what this process looks like, you can view this video - https://youtu.be/8T8RFwI_dZ0 .

Written Questions will be manually graded by Course Facilitators no later than one week after a problem set's 'on-time' due date. Problem sets turned in late may be graded slightly later. To view an example of what the written submission process looks like, you can view this video - https://youtu.be/eEn826KNUqw .

Note on Final Projects

In the adaption from the CS229 graduate course to this professional course, the original CS229 course has been split in half. The final project component is included in the second course (XCS229ii: Machine Learning Strategy and Intro to Reinforcement Learning). The first cohort of XCS229ii will launch in March 2021, following the completion of this course.

Late Problem Sets and One-time Penalty Waiver

Late problem sets are assessed a penalty of **one point per day late, up to a maximum of five days late at which point the submission link will close**.

We understand that personal or professional events may cause you to miss a deadline on a problem set. Each student can use a **one-time penalty waiver on any late problem set, which will not be assessed the late penalty.** The extension cannot be split into smaller parts (e.g., you cannot use two days on Problem Set 3 and three days on Problem Set 4.). In order to use your extension, contact your Course Facilitator or SCPD staff.

Passing the Course and Earning the Certificate

In order to earn the Certificate of Achievement associated with this course, you must complete problem sets with a total cumulative score of 70% or higher. Once you have successfully completed the course and the post-

class survey, a digital Record of Completion will be emailed to you and the Certificate of Achievement will be mailed in a Stanford holder in approximately four weeks. If you are interested in calculating your progress along the way, it may be helpful to know:

- There are a total of 200 base points in the course (meaning 140 to achieve 70%)
- You can earn up to 40 points on each problem set.

Videos and Slides

As noted, this course utilizes content originally delivered in the CS229 graduate course. A few things you will notice about this adaptation process:

- At times you will hear instructors refer to the final project or poster session. As noted above these have been removed from the current version of XCS229i and you need not worry about the reference.
- Instructors may refer to “Week 1”, “Week 2”, “Week n” of the course – these references can be ignored.
- In a few specific cases you may see names and/or faces blurred. In general this is usually due to guidelines regarding student privacy.

Course Facilitators, Support, and Guidelines

You have a wide range of support available to you throughout the course. You will be assigned and receive contact information for an individual Course Facilitator (CF) who will act as your primary point of contact. Below is a summary of the available resources and course support:

Office Hours

Your CF will be in touch with availability and scheduling logistics for video conference office hours. Office hours may be conducted using the Zoom conference service or via Slack video (more information below on the course Slack workspace). Your CF will provide further information on how they will schedule and run office hour sessions.

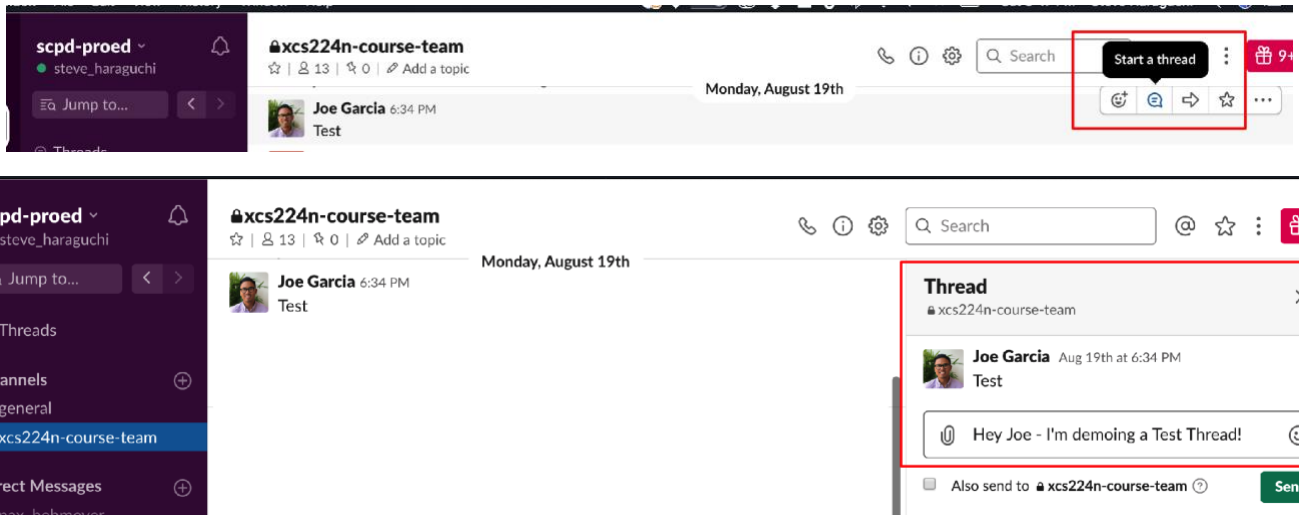
Email

Your CF will also be available to answer questions via email – a Stanford contact address will be provided when you are first connected to your CF.

Slack Workspace – Usage and Guidelines

In addition to the individual and small group support provided by CFs, Slack will be a place where questions may be posed to the entire community (this is the fastest way to get an answer!). **In order to keep the Slack workspace readable, searchable, and useful to all, please follow the following guidelines:**

Reply in Threads to Keep Conversation Organized – When you are replying to a post or joining a conversation, respond by starting or joining a threaded conversation, rather than responding in the full flow of the standard timeline. See below for an example of how to respond in a threaded conversation to Joe's test message:



Use Multi-line Messages – Even if messages are threaded, you will soon see that Slack becomes unmanageable unless people use **single, multi-line messages instead of multiple, single-line messages**. Especially for mobile Slack users, it gets out of control!

Rather than the following:
"Hey all I have a question" [RETURN] ← Creates new message
"I am a little confused about the quiz" [RETURN] ← Creates new message
"I'm getting F for Question 40, but it seems like T is better" [RETURN] ← Creates new message
---
Instead, try this!
"Hey all I have a question" [SHIFT+RETURN] ← Creates new line in SAME message
"I am a little confused about the quiz" [SHIFT+RETURN] ← Creates new line in SAME message
"I'm getting F for Question 40, but it seems like T is better" [SHIFT+RETURN] ← Creates new line in SAME message
[RETURN] ← Posts message

Note on Code Questions and Debugging

While the course team is here to help and support your experience, it is ultimately your responsibility to write, test, and de-bug your own code. CFs may view and provide guidance on your work, however they will not send you exact answers on what to insert into your problem sets. Additionally, before reaching out to a CF or Slack for help, it's expected that you have taken the reasonable steps of reading and performing an analysis yourself. This policy is meant to ensure that you leave the course having mastered the material and enable CFs to focus attention on questions where their guidance is most impactful.

Drop/Transfer Policy

You may drop this course for a full refund up until April 11th, 2021 – the day before the course starts. Once the course has begun, if you request to drop the course by Friday at 5:00pm PST on the third week of the cohort (**April 30th, 2020**) you will be reimbursed 100% of your tuition minus a drop fee of $100. Beyond the third week of the course, tuition refunds are not granted. Up until **April 30th** you may also request to transfer your enrollment to a future cohort of XCS229i or another course in the AI Professional Program, also for a transfer fee of $100.

Questions and Contacts? Email xcs229i-staff@stanford.edu