# XCS229: Machine Learning

# Syllabus and Course Information

## Welcome

Welcome to XCS229: Machine Learning! This professional course is based on graduate-level material from Stanford's on-campus course CS229, adapted for a professional certificate format. 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 implementing content covered in the course lectures, 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 11th) at 12:00pm Pacific Time**. Course problem sets will be released as indicated in the course schedule below, without exception. Maintaining the assignment schedule enables Course Facilitators to be most effective in providing support and answering questions on subject matter throughout the course.

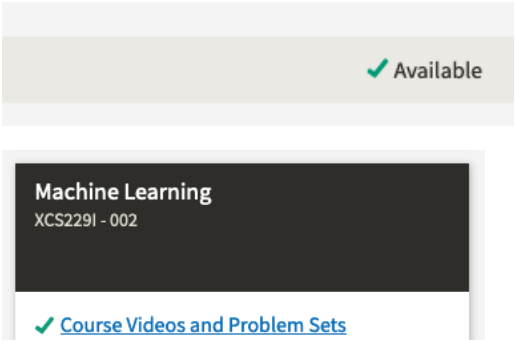
# Getting Started (On April 11)

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.

## Accessing Your Course

1. On **April 11th after 12pm Pacific Time**, log in to the mystanfordconnection account you used when applying for the XCS229 course.
2. XCS229: Machine Learning will be visible as a live course. Click the link titled "Course Videos and Problem Sets" to enter our learning management system.




## 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 8th**.

**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 8th** should proceed directly to http://xcs229-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 **"XCS229-Spring-2022"** at your address on file with SCPD. If you'd prefer to receive a GitHub invitation at a different address, just let us know at xcs229-staff@stanford.edu.

**Please, accept the GitHub invitation as soon as possible, as it expires within 7 days. Once your invite expires, ask the team to resend it:** xcs229-staff@stanford.edu.

**You will need to accept the invitation and be logged into your GitHub account to view course assignments and code.**

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

## Course Calendar and Potential Pacing Guide

Below is a ***potential*** pacing guide if you are interested in watching all videos within a 10-week window. However, you are free to view the videos at any pace you'd like.

| Week | Suggested Videos | **Optional** Live Sessions (will also be recorded) | Problem Sets |
|---|---|---|---|
| Week 1 & 2 (Apr 11 - 24) | **Module 1: Supervised Learning**<br>*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 11<br>PS 2 Released – Monday, Apr 11<br>PS 1 Due – Sunday, Apr 24 |
| Week 3 & 4 (Apr 25 - May 8) | **Module 1:**<br>*Perceptron, Exponential Family, GLM, Softmax Regression, GDA, Generative & Discriminative Analysis* | **Problem Set 1 Solution Walk-Through**<br>Zoom - Times/Dates TBD | PS 3 Released – Friday, Apr 29<br>PS 2 Due – Sunday, May 8 |
| Week 5 & 6 (May 9 - 22) | **Module 1:**<br>*Naive Bayes, Laplace Smoothing & Event Models, SVM, Kernels, Decision Trees* | **Problem Set 2 Solution Walk-Through**<br>Zoom - Times/Dates TBD | PS 4 Released – Friday, May 13<br>PS 3 Due – Sunday, May 22 |
| Week 7 & 8 (May 23 - Jun 5) | **Module 1 & 2 & 3**<br>*Ensemble Methods, Neural Networks, K-means Clustering, Mixture of Gaussians* | **Problem Set 3 Solution Walk-Through**<br>Zoom - Times/Dates TBD | PS 5 Released – Friday, May 27<br>PS 4 Due – Sunday, Jun 5 |
| Week 9 & 10 (Jun 6 - 19) | **Module 3**<br>*Expectation-Maximization Algorithm, Factor Analysis, Learning Theory, Independent Component Analysis* | **Problem Set 4 & 5 Solutions Walk-Through**<br>Zoom - Times/Dates TBD | PS 5 Due – Sunday, Jun 19 |

## 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?
3. Logistic Regression: Training stability

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. In the last problem you will dive into how the training stability of logistic regression differs on diverse datasets and best debugging techniques to enhance machine learning algorithms like logistic regression.

**Problem Set 2**

1. Linear Classifiers (logistic regression and GDA)
2. Incomplete, Positive-Only Labels

In this problem set, you will use a discriminative linear classifier, logistic regression to 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 and the hand-tuning needed for a diverse set of classification tasks.

**Problem Set 3**

1. Poisson Regression
2. Constructing Kernels
3. Kernelizing the Perceptron

In the first problem, you will construct the Poisson Regression GLM. 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. In the second problem, you will learn to construct kernels and gain a discerning eye for the general requirements for

a function to be a valid kernel. In the final problem, you will apply the kernel definitions to the perceptron model to create non-linear decision boundaries for your data.

**Problem Set 4**

1. KL Divergence and Maximum Likelihood
2. Neural Networks: MNIST Image Classification
3. Spam Classification

In the first problem you will use the Kullback-Leiber (KL) divergence, a measure of how one probability distribution differs from another, to practice simple manipulations over discrete distributions and connect it back to Maximum Likelihood estimation. In the second problem, you will create a neural network to solve the classic data science problem of classifying the MNIST dataset. In the last problem, you will use the naive-Bayes algorithm and a SVM to build a spam classifier.

**Problem Set 5**

1. PCA
2. Semi-supervised EM
3. Independent Component Analysis

In the first problem, you will apply machine learning theory to the principle component analysis (PCA) algorithm to give another interpretation to PCA. 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. For the final problem, you will implement ICA to find a solution to the cocktail party problem.

## 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.

We encourage students to form study groups. Students may discuss and work on homework problem sets in groups. However, each student must write down the solutions independently, and without referring to written notes from the joint session. **In other words, each student must understand the solution well enough in order to reconstruct it by him/herself.** In addition, each student should write on the problem set the set of people with whom s/he collaborated. Further, because we occasionally reuse problem set questions from previous years, we expect students not to copy, refer to, or look at the solutions in preparing their answers. It is an honor code violation to intentionally refer to a previous year's solutions.

After completing this course, you are welcome to share your experience and credential with others; However, **it is considered a violation of the honor code to share problem set solutions including on public platforms such as GitHub**. Faculty in the computer science department have strongly encouraged us to refrain from posting solutions for problem sets, thus we ask that you DO NOT share the exact code.

## 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 XCS229 professional course, the final project has been removed.

## Late Problem Sets and One-time Penalty Waiver

We understand that personal or professional events may cause you to miss a regular deadline on an assignment. All assignments can be turned in **up to five days late and are assessed a penalty of -1 points per late day.** After five days, the submission link will close, and entries will no longer be accepted.

Each student has an option to use a **one-time penalty waiver to remove late penalty points from one assignment**. Note, that the waiver does **NOT** serve as an extension, and **cannot** be split into smaller parts (e.g., you cannot use two days on Problem Set 3 and three days on Problem Set 1). In order to use your penalty waiver, contact your Course Facilitator or SCPD staff at xcs229-staff@stanford.edu

## 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 badge for course completion** will be emailed to you, which is **the** official certified Stanford document generated and held by Stanford Online.

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%)

| Deliverable | Points |
|---|---|
| A1 (Coding)<br>A1 (Written) | 13<br>20 |
| A2 (Coding)<br>A2 (Written) | 20<br>20 |
| A3 (Coding)<br>A3 (Written) | 16<br>20 |
| A4 (Coding)<br>A4 (Written) | 30<br>15 |
| A5 (Coding)<br>A5 (Written) | 21<br>25 |
| **Total Available** | **200** |
| **Minimum Passing Total** | **140 (70 %)** |

**Note, that along with moving towards more flexible and learner-friendly digital badges, SCPD is no longer issuing hard copies of course certificates to promote *Stanford green sustainable initiative*.

## 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 XCS229 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.

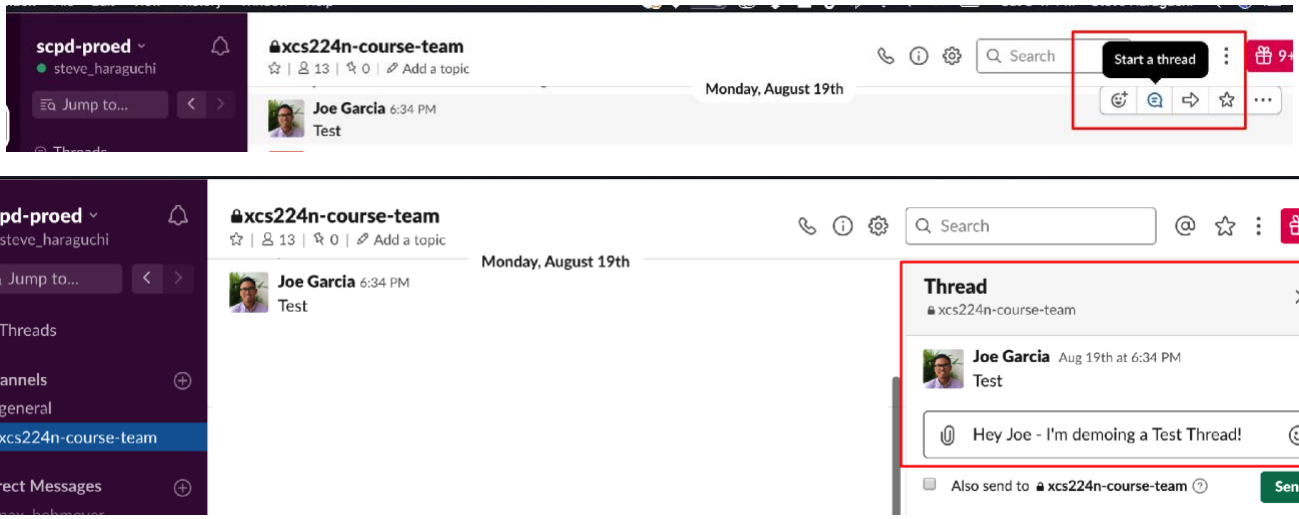
## Email

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.

## Note on Networking

One of the benefits of this course is to be able to network with other course participants and create study groups. We encourage this kind of interaction and want to make sure that it is a positive experience. It is imperative that no course participant is made to feel uncomfortable or their ability to learn or otherwise benefit from the course is impeded by the actions of another participant. Please use good judgment. Keep interactions professional and focused on coursework or career networking. Avoid using offensive language and respect your colleagues' preferences regarding direct messaging. Please respect and uphold the rights and dignity of others regardless of race, color, national or ethnic origin, sex, age, disability, religion, sexual orientation, gender identity, or socio-economic status. Our team is always available either here in Slack or via email, so please feel free to reach out to us if you have any questions, concerns, or if any situation arises.

You can review SCPD's terms of service here, including rules for online conduct.

## Drop/Transfer Policy

You may drop this course for a full refund up until **April 11th, 2022** – the day 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 29th, 2022)** 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. Before **April 29th, 2022** you may also request a transfer to a different course in the AI Professional Program, also for a $100 transfer fee. To request a drop or transfer, send an email to xcs229-staff@stanford.edu

**Please note that drop refunds require longer processing time, as well as a few additional steps **if** the payment was received by SCPD 6 months ago or earlier.

## Questions

For course-specific questions or concerns (content, assignments, CF support), please contact your designated Course Facilitator.

For other course related questions, email xcs229-staff@stanford.edu