UC San Diego

EXTENDED STUDIES

UCSD Extended Studies Machine Learning Engineering Bootcamp

Syllabus & Course Overview

Introduction

Machine learning engineering is concerned with the creation, development, and deployment of intelligent software that learns from Big Data and provides cutting-edge business insights.

Today, some of the most exciting and revolutionary technology in the world is powered by machine learning: including software that detects cancer from image files, steers driverless cars, and recommends products that millions will buy. The mainstream adoption of AI and ML-powered products, coupled with ever-bigger bets by tech and tech-adjacent companies, has led to an AI talent shortage that many experts predict won't be alleviated for years. Consider this:

- The global machine learning (ML) market was valued at $1.58 billion in 2017 and is expected to grow to $20.83 billion by 2024 (Source: Forbes)
- From 2015 to 2018, the number of AI-related job postings on Indeed increased by 119 percent (Source: Indeed)
- Machine learning is the most in-demand AI skill (Source: Indeed)

All of that is driving compensation sky high. Machine learning engineers earn an average salary between $125,000 and $175,000. AI engineer salaries at the ten highest-paying companies start above $200,000 a year.

The UCSD Machine Learning Bootcamp is focused on outfitting students with the skills they need to bridge the gaps in the current job market. This Machine Learning Bootcamp is an intensive 6-month online program that will prepare you for a role as a Machine Learning Engineer (MLE).

Our curriculum is rigorous and deeply technical, teaching you the foundations of machine learning and deep learning. We don’t just focus on concepts; we focus on the creation and deployment of applications that apply these concepts in a hands-on, end-to-end manner that mirrors work you’ll be doing on the job.

We don’t just teach you about machine learning, you do machine learning! Of the 450+ hours of core and optional work it takes to complete this course, 200 hours are dedicated to projects and hands-on exercises. You’ll build and deploy large-scale ML systems — with 1:1 guidance from your personal mentor, an experienced machine learning engineer currently working in the industry.

Who's It For?

University of California San Diego Extended Studies' Machine Learning Bootcamp is an advanced course for technically minded students who want to develop their machine learning skills to pursue a career as a machine learning engineer. It is open to students who are working as software engineers or data scientists, students who have undergraduate degrees in computer science, physics, computational mathematics, statistics, or a similar field. The course is also open to self-taught programmers who display a high degree of technical savvy.

How It Works

1. **Cost and Schedule:** The course costs $10,340, paid upfront (other payment options available). It is fully online and features a flexible pace so that you can study anytime, anywhere, even if you have a full-time job. Students can complete the course in 6 months if they dedicate approximately 15 hours of work per week. You're welcomed to complete in less or more time — you pay only for the months you're Enrolled.
2. **Enrollment:** We have monthly cohorts — enrollment opens a few weeks before each class is set to begin. If enrollments aren't open, you can sign up for the waitlist on the course page.
3. **Mentor-matching process:** Once you enroll, you'll be asked to fill out a profile, which includes questions about your background, your availability during the week, and the skills you want to develop. Your student advisor will use this information to match you with a mentor who suits your specific needs.
4. **A curriculum curated by experts:** We believe that diverse perspectives lead to better learning outcomes. Our online curriculum is packed with quality AI and machine learning resources, some of which are handpicked from around the web by industry experts and others that are created in-house. Our instructional designers then build these resources into a curriculum that includes projects and quizzes.
5. **Your own support team:** including a student advisor, mentor, community manager, and career coach.
    a. Your student advisor will match you with a mentor, help you prepare for the course, and answer your general questions.
    b. You'll have 1-on-1 calls with your mentor each week. They'll provide feedback on projects, answer questions about the curriculum, and give you career advice and industry insight.

    c. Your community manager can answer questions about the curriculum and machine learning engineering industry.
    d. Your career coach will help you during your job hunt and can give you tips about how to network, create a strong machine learning engineer resume, and more.
6. **The program community:** While online learning may sound isolating, it's important to remember that you have a whole community learning alongside you. You'll get access to this community so you can share triumphs and trials, get feedback, and attend weekly live Office Hours.
7. **Career services:** In addition to learning about machine learning, you'll have the option to complete optional units that will guide you through your job search.
8. **Certification and alumni status:** Once you finish the course material and submit your capstone project, you'll get a certificate of completion and access to the UCSD Extended Studies Alumni community. **You will also receive continuing education unit (CEU) credits that can be applied to other UCSD Extended Studies credentials.**

Syllabus

This curriculum includes 450+ hours of core and optional content that will teach you the foundations of machine learning and deep learning — and how to implement them at scale. You'll learn the best practices for working with data, build and scale prototypes of various ML models, deploy these models to production, advanced specializations like NLP and Computer Vision, all while working on an epic Capstone Project of your choice.

Modules Include:

Machine Learning Models

We'll teach you the most in-demand machine learning models and algorithms you'll need to know to succeed as an MLE. For each model, you will learn how it works conceptually first, then the applied mathematics necessary to implement it, and finally you will get experience training and testing the models. We'll walk you through the best practices for predictive optimization, like hyperparameter tuning, and how to evaluate your performance. You'll learn how to pick the right model for the challenge you are facing, and critically, how to implement and deploy these models at scale.

Topics Covered:

1. Regression modeling with linear and logistical regression

2. Classification modeling with naive bayes, k-nearest neighbor, and support vector machines
3. Decision tree models with random forest and the accompanying boosting algorithms such as XGBoost and CatBoost
4. Deep learning techniques like deep, convolutional, and recurrent neural networks, and generative adversarial networks
5. Anomaly detection modeling with isolated forests, PCA, and K-Means clustering
6. Recommendation systems and time series prediction models
7. Model selection, evaluation, and interpretation concepts like regularization, the Curse of Dimensionality, and cross-validation
8. Tools: scikit-learn, SparkML, AutoML systems

A Stack For Machine Learning Engineering

Throughout this course, you'll be introduced to a variety of tools and libraries that are used in both data science and machine learning. These include everything from ML libraries to deployment tools. There will also be refreshers on software engineering best practices and foundational math concepts that every ML Engineer should know.

Topics Covered:

1. Python Data Science Tools includes pandas, scikit-learn, Keras, TensorFlow
2. Machine learning engineering tools including Spark/PySpark, TensorFlow, Luigi, Docker, Hadoop, AWS, and Fast.ai
3. Software engineering tools including continuous integration, version control with Git, logging, testing, and debugging
4. Deployment tools like Paperspace, FastAPI, AWS, and Algorithmia
5. Working With Data

Data, The Fuel of Machine Learning

A critical part of every machine learning engineer's job is collecting, cleaning, processing, and transforming data. Without quality data, you can't get quality insights. You'll learn the best practices and tools for working with data at scale and how to transform a messy, sparse dataset into something worthy of modeling.

Topics Covered:

1. Collecting data from APIs, RSSs, and web scraping
2. Cleaning and transforming data for ML systems at scale, including tools for automatic transformation
3. Working with large data sets in SQL and NoSQL databases
4. Tools such as pandas, Spark, Dask, SQL, Spark SQL, and ScrappingHub

---

Essential Mathematics and Statistics

Throughout the course, you'll learn about the fundamental mathematical and statistical concepts that make up the core of the field of machine learning, including calculus and linear algebra. You'll learn statistics and probability concepts such as conditional probability, independence, and Bayes theorem. You'll also learn about Probabilistic Programming, a new set of techniques used to quickly assess risk in systems in a highly interpretable manner.

Topics Covered:

1. Essential mathematics like calculus and linear algebra
2. Stats concepts like conditional probability, independence, and Bayes Theorem
3. Probabilistic programming

Machine Learning Models At Scale and In Production

Machine learning at scale and in production is an entirely different beast than training a model in Jupyter notebook. When you're working at scale, there are a host of problems that can disrupt your model and its performance. We'll teach you about the best practices for surmounting these challenges, how to write production-level code, as well as ensuring that you are getting quality data fed into your model.

Topics Covered:

1. Creating reliable and reproducible data pipelines to ensure your model is well fueled

2. Cloud-based services provided by AWS, Microsoft Azure, and Google
3. Using Dask and pandas to scale large datasets
4. Using SparkML to scale an ML model, debugging and monitoring Spark ML applications and pipelines
5. The machine learning life cycle and challenges that can occur when integrating your model into an application
6. Machine Learning API tools like Swagger and Postman

Deep Learning

Deep learning is a school of machine learning that involves the training of self-generating neural networks, which take their inspiration from the inner workings of the human brain. This advanced machine learning technique powers many of today's most cutting edge applications, including generating photorealistic faces of people who have never lived, machine translation, self-driving cars, speech recognition, and more. Deep learning models become more accurate when they are fed more data, so they are excellent for many business problems.

Topics Covered:

1. Overview of Neural Networks, backpropagation, and foundational techniques like stochastic gradient descent
2. Principles of Deep Neural Networks
3. Common Deep Neural Network configurations e.g. RNNs, CNNs, MLPs, LSTMs
4. Generative Deep Learning and GANs
5. Linear algebra and calculus necessary for these models
6. Engineering Frameworks like Keras, TensorFlow, PyTorch, Fast.ai, and CuPy

---

Natural Language Processing

NLP uses techniques from computer science, linguistics, and machine learning to process human language, typically in the form of unstructured text. You'll learn the best practices for working with text data, how to clean and process it, and how to extract meaningful insights like sentiment analysis from text sources and conversations. We'll

walk you through a detailed case study to solve a real NLP problem using Deep Learning.

Topics Covered:

1. How to work with text and natural language data
2. NLP in Python, using common libraries such as NLTK, Flair, and spaCy
3. Representing language: BOW, TF-IDF, word embedding models (word2vec, GloVe, FastText, and StarSpace)
4. Deep Learning and Transfer Learning techniques for NLP
5. Chatbots and other modern NLP applications

Computer Vision and Image Processing

The field of Computer Vision, which focuses on image recognition and the creation of unique images, is rapidly evolving because of the wealth of image data proliferated through social media and other online sources. Computer Vision is a highly innovative and cutting edge field of machine learning and artificial intelligence, there is a world of opportunities for an engineer specializing in CV. Computer vision and image processing concepts will be spread across two units — one that dives into the theory behind these concepts and another that works through a hands-on tutorial that will help you put into practice everything you've learned.

Topics Covered:

1. Foundations of computer vision and image processing, including an introduction to OpenCV and how to use neural networks for image processing
2. Image clustering and classification with K-means, multitask classifiers, and GANs
3. Object detection and image segmentation with techniques like Single Shot Detectors and YOLO Detection
4. Applications and trends in computer vision

Deploying ML Systems to Production

Deploying machine learning models to production in an at-scale environment is perhaps

the most difficult part of a MLE's job. It can take over a month to properly deploy a model, and most other machine learning courses do not focus on deployment, which is a much-desired skill in the workplace. We will teach you a wide range of tools, identify common pitfalls, and you will get hands-on experience deploying models. If you know nothing about deployment, have no fear! We will have optional resources to help you get up to speed with how it works!

Topics Covered:

1. Common tools and techniques to build large-scale AI applications
2. Tools for building and deploying quality APIs like: Swagger, Postman, FastAPI, and Paperspace
3. Productionizing models with CI and CD
4. Packaging your model into an interactive product like an app or website with tools like Streamlit, TensorFlow.js, and TensorFlow Lite
5. Tools like PySpark, PyTorch, and Spark for model production

---

Capstone Project

The Capstone Project is a mandatory part of our curriculum. This course has one capstone project that has been split up into two phases. Using a combination of tools and techniques that you've learned, you'll build a realistic, complete ML or DL application.

Work on your capstone project will involve:

Phase One: Building a working prototype

1. Step One: Pick your initial project ideas.
2. Step Two: Write your project proposal.
3. Step Three: Collect your data.
4. Step Four: Data wrangling and exploration.
5. Step Five: Create a machine learning or deep learning prototype.
6. Step Six: Scale your prototype.

Phase Two: Deploying your prototype to production.

1. Step One: Create a deployment architecture.
2. Step Two: Run your code end-to-end with testing.
3. Step Three: Deploy your application to production.

4. Extra Credit Step: Build a web interface to your application

UCSD will provide you free access to a cloud-based engineering environment, which will support all of the standard tools and libraries.

Career Resources (35+ hours)

Our optional career material is designed to help you create a tailored job search strategy based on your background and goals. You'll learn how to craft a resume that stands out from the pack, evaluate companies and roles, ace interviews, and negotiate the best possible salary. Your career coach will be with you every step of the way, offering feedback and providing personalized tips based on your goals.

Topics Covered:

1. Job search strategies that top candidates use
2. How to build your network and effectively use it to land interviews
3. Create a high-quality resume, LinkedIn profile and cover letter
4. Interview coaching and practice, including mock interviews for both technical and non-technical topics
5. Negotiation success tips

Ready for the next step?
Learn more and apply here.

Email us at **ucsd@career-bootcamp.com** with any questions.

Powered by Springboard