UDACITY

Intro to Machine Learning with PyTorch

Overview

This ultimate goal of the Intro to Machine Learning Nanodegree program is to help students learn machine learning techniques such as data transformation and algorithms that can find patterns in data and apply machine learning algorithms to tasks of their own design.

A graduate of this program will be able to:

- Use Python and SQL to access and analyze data from several different data sources.
- Build predictive models using a variety of unsupervised and supervised machine learning techniques.
- Perform feature engineering to improve the performance of machine learning models.
- Optimize, tune, and improve algorithms according to specific metrics like accuracy and speed.
- Compare the performances of learned models using suitable metrics.

One of our main goals at Udacity is to help you create a job-ready portfolio of completed projects. Building a project is one of the best ways to test the skills you've acquired and to demonstrate your newfound abilities to future employers or colleagues. Throughout this Nanodegree program, you'll have the
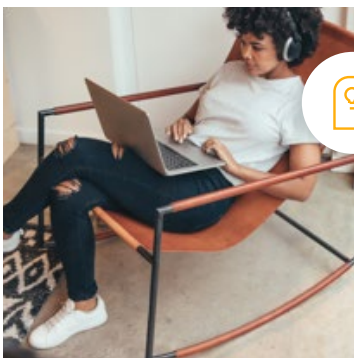
IN COLLABORATION WITH

kaggle   aws

Estimated Time: 3 Months at 10hrs/week
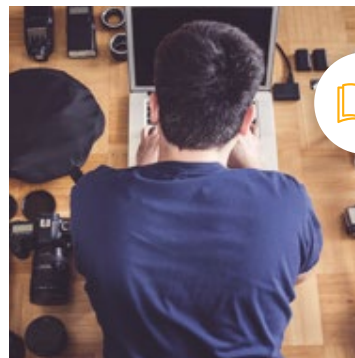
Prerequisites: Python

Flexible Learning: Self-paced, so you can learn on the schedule that works best for you

Technical Mentor Support: Our knowledgeable mentors guide your learning and are focused on answering your questions, motivating you and keeping you on track

opportunity to prove your skills by building the following projects:

- Finding Donors for CharityML
- Apply supervised learning techniques on data collected for the US census to help CharityML (a fictitious charity organization) identify groups of people that are most likely to donate to their cause.
- Create Your Own Image Classifier
- Define and train a neural network in PyTorch that learns to classify images; going from image data exploration to network training and evaluation.
- Identify Customer Segments with Arvato
- Study a real dataset of customers for a company, and apply several unsupervised learning techniques in order to segment customers into similar groups and extract information that may be used for marketing or product improvement.

In the sections below, you'll find detailed descriptions of each project along with the course material that presents the skills required to complete the project.

Course 1:   Supervised Learning

One of our main goals at Udacity is to help you create a job-ready portfolio of completed projects. Building a project is one of the best ways to test the skills you've acquired and to demonstrate your newfound abilities to future employers or colleagues. Throughout this Nanodegree program, you'll have the opportunity to prove your skills by building the following projects:

• Finding Donors for CharityML
• Apply supervised learning techniques on data collected for the US census to help CharityML (a fictitious charity organization) identify groups of people that are most likely to donate to their cause.
• Create Your Own Image Classifier
• Define and train a neural network in PyTorch that learns to classify images; going from image data exploration to network training and evaluation.
• Identify Customer Segments with Arvato
• Study a real dataset of customers for a company, and apply several unsupervised learning techniques in order to segment customers into similar groups and extract information that may be used for marketing or product improvement.

In the sections below, you'll find detailed descriptions of each project along with the course material that presents the skills required to complete the project.

Course Project
Find Donors for CharityML

CharityML is a fictitious charity organization located in the heart of Silicon Valley that was established to provide financial support for people eager to learn machine learning. To expand their potential donor base, CharityML has decided to send letters to residents of California, but to only those most likely to donate to the charity. Your goal will be to evaluate and optimize several different supervised learning algorithms to determine which algorithm will provide the highest donation yield while under some marketing constraints.
• Supervised learning
• Model evaluation and comparison
• Tuning models according to constraints

	LEARNING OUTCOMES	
LESSON ONE	**Regression**	• Learn the difference between Regression and Classification • Train a Linear Regression model to predict values • Learn to predict states using Logistic Regression

LESSON TWO	**Perceptron Algorithms**	• Learn the definition of a perceptron as a building block for neural networks, and the perceptron algorithm for classification.
LESSON THREE	**Decision Trees**	• Train Decision Trees to predict states • Use Entropy to build decision trees, recursively
LESSON FOUR	**Naive Bayes'**	• Learn Bayes' rule, and apply it to predict cases of spam messages using the Naive Bayes algorithm. • Train models using Bayesian Learning • Complete an exercise that uses Bayesian Learning for natural language processing
LESSON FIVE	**Support Vector Machines**	• Learn to train a Support Vector Machines to separate data, linearly • Use Kernel Methods in order to train SVMs on data that is not linearly separable
LESSON SIX	**Ensemble of Learners**	• Build data visualizations for quantitative and categorical data. • Create pie, bar, line, scatter, histogram, and boxplot charts. • Build professional presentations.
LESSON SEVEN	**Evaluation Metrics**	• Learn about different metrics to measure model success. • Calculate accuracy, precision, and recall to measure the performance of your models.
LESSON EIGHT	**Training and Tuning Models**	• Train and test models with Scikit-learn. • Choose the best model using evaluation techniques like cross-validation and grid search.

Course 2: Neural Networks

Implementing an image classification application using a deep neural network. This application will train a deep learning model on a dataset of images. It will then use the trained model to classify new images. You will develop your code in a Jupyter notebook to ensure your implementation works well.

Course Project
Create Your Own Image Classifier

As a machine learning engineer at a fictional self-driving car startup, you have been asked to help decide whether to build or buy an object detection algorithm for objects that may be on the side of the road. A company, Detectocorp, claims an 80% accuracy rate on the CIFAR-10 dataset, a benchmark used to evaluate the state of the art for computer vision systems.

You will try your hand at training a neural network to recognize objects in images and evaluate the model's performance compared to Detectocorp's model.

	LEARNING OUTCOMES	
LESSON ONE	**Introduction to Neural Networks**	• Learn the foundations of deep learning and neural networks. • Implement gradient descent and backpropagation in Python.
LESSON TWO	**Implementing Gradient Descent**	• Implement gradient descent using NumPy matrix multiplication.
LESSON THREE	**Training Neural Networks**	• Learn several techniques to effectively train a neural network. • Prevent overfitting of training data and learn best practices for minimizing the error of a network.
LESSON FOUR	**Deep Learning with PyTorch**	• Learn how to use PyTorch for building deep learning models.

Course 3: Unsupervised Learning

Course Project
Creating Customer Segments

In this project, you will apply unsupervised learning techniques on product spending data collected for customers of a wholesale distributor in Lisbon, Portugal to identify customer segments hidden in the data. You will first explore and pre-process the data by scaling each product category and then identifying (and removing) unwanted outliers. With the cleaned customer spending data, you will apply PCA transformations to the data and implement clustering algorithms to segment the transformed customer data. Finally, you will compare the segmentation found with an additional labeling and consider ways this information could assist the wholesale distributor with future service changes.

Key Skills Demonstrated:
- Data cleaning
- Dimensionality reduction with PCA
- Unsupervised clustering

|	LEARNING OUTCOMES	
LESSON ONE	**Clustering**	• Learn the basics of clustering data • Cluster data with the K-means algorithm
LESSON TWO	**Hierarchical and Density-Based Clustering**	• Cluster data with Single Linkage Clustering. • Cluster data with DBSCAN, a clustering method that captures the insight that clusters are dense group of points
LESSON THREE	**Gaussian Mixture Models**	• Cluster data with Gaussian Mixture Models • Optimize Gaussian Mixture Models with and Expectation Maximization
LESSON FOUR	**Dimensionality Reduction**	• Reduce the dimensionality of the data using Principal • Component Analysis and Independent Component Analysis

Our Classroom Experience
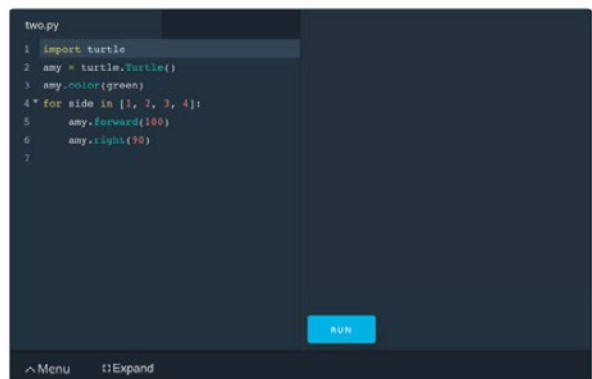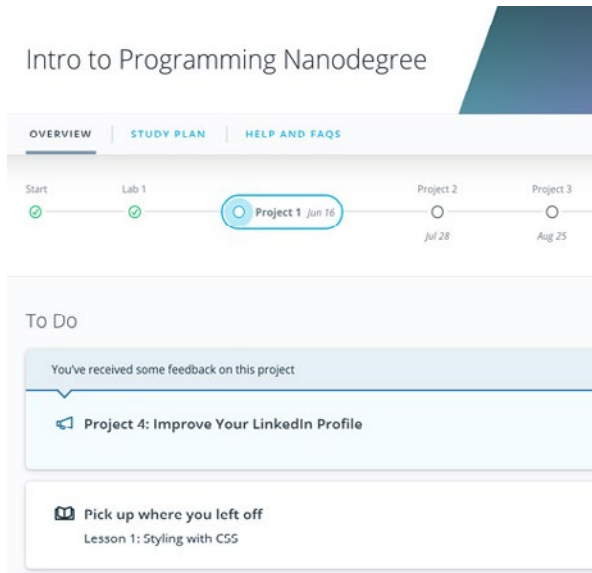
REAL-WORLD PROJECTS
Build your skills through industry-relevant projects. Get personalized feedback from our network of 900+ project reviewers. Our simple interface makes it easy to submit your projects as often as you need and receive unlimited feedback on your work.

KNOWLEDGE
Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students, connect with technical mentors, and discover in real-time how to solve the challenges that you encounter.

WORKSPACES
See your code in action. Check the output and quality of your code by running them on workspaces that are a part of our classroom.

QUIZZES
Check your understanding of concepts learned in the program by answering simple and auto-graded quizzes. Easily go back to the lessons to brush up on concepts anytime you get an answer wrong.

CUSTOM STUDY PLANS
Create a custom study plan to suit your personal needs and use this plan to keep track of your progress toward your goal.

PROGRESS TRACKER
Stay on track to complete your Nanodegree program with useful milestone reminders.

Learn with the Best

Luis Serrano

INSTRUCTOR

Luis was formerly a Machine Learning Engineer at Google. He holds a PhD in mathematics from the University of Michigan, and a Postdoctoral Fellowship at the University of Quebec at Montreal.

Cezanne Camacho

CURRICULUM LEAD

Cezanne is a machine learning educator with a Masters in Electrical Engineering from Stanford University. As a former researcher in genomics and biomedical imaging, she's applied machine learning to medical diagnostic applications.

Dan Romuald Mbanga

INSTRUCTOR

Dan leads Amazon AI's Business Development efforts for Machine Learning Services. Day to day, he works with customers—from startups to enterprises—to ensure they are successful at building and deploying models on Amazon SageMaker.

Mat Leonard

INSTRUCTOR

Mat is a former physicist, research neuroscientist, and data scientist. He did his PhD and Postdoctoral Fellowship at the University of California, Berkeley.
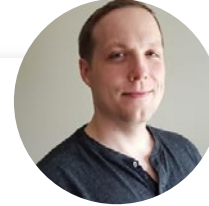
Learn with the Best

Jennifer Staab

INSTRUCTOR

Jennifer has a PhD in Computer Science and a Masters in Biostatistics; she was a professor at Florida Polytechnic University. She previously worked at RTI International and United Therapeutics as a statistician and computer scientist.

Sean Carrell

INSTRUCTOR

Sean Carrell is a former research mathematician specializing in Algebraic Combinatorics. He completed his PhD and Postdoctoral Fellowship at the University of Waterloo, Canada.

Josh Bernhard

DATA SCIENTIST

Josh has been sharing his passion for data for nearly a decade at all levels of university, and as Lead Data Science Instructor at Galvanize. He's used data science for work ranging from cancer research to process automation.

Mat Leonard

INSTRUCTOR

Jay has a degree in computer science, loves visualizing machine learning concepts, and is the Investment Principal at STV, a $500 million venture capital fund focused on high-technology startups.
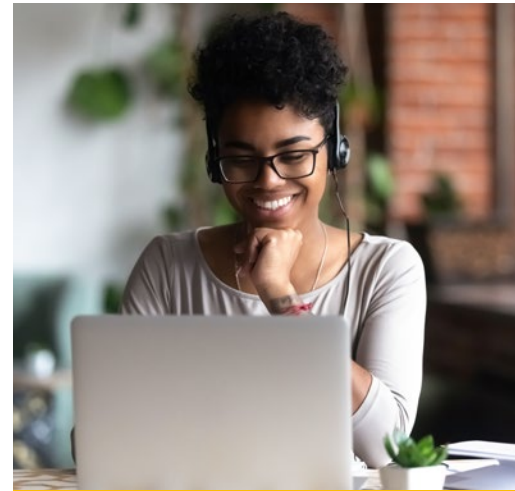
All Our Nanodegree Programs Include:

EXPERIENCED PROJECT REVIEWERS

REVIEWER SERVICES

- Personalized feedback & line by line code reviews
- 1600+ Reviewers with a 4.85/5 average rating
- 3 hour average project review turnaround time
- Unlimited submissions and feedback loops
- Practical tips and industry best practices
- Additional suggested resources to improve

TECHNICAL MENTOR SUPPORT

MENTORSHIP SERVICES

- Questions answered quickly by our team of technical mentors
- 1000+ Mentors with a 4.7/5 average rating
- Support for all your technical questions

PERSONAL CAREER SERVICES

CAREER SUPPORT

- Github portfolio review
- LinkedIn profile optimization

Frequently Asked Questions

PROGRAM OVERVIEW

WHY SHOULD I ENROLL?
Machine learning is changing countless industries, from health care to finance to market predictions. Currently, the demand for machine learning engineers far exceeds the supply. In this program, you'll apply machine learning techniques to a variety of real-world tasks, such as customer segmentation and image classification. This program is designed to teach you foundational machine learning skills that data scientists and machine learning engineers use day-to-day.

WHAT JOBS WILL THIS PROGRAM PREPARE ME FOR?
This program emphasizes practical coding skills that demonstrate your ability to apply machine learning techniques to a variety of business and research tasks. It is designed for people who are new to machine learning and want to build foundational skills in machine learning algorithms and techniques to either advance within their current field or position themselves to learn more advanced skills for a career transition.

HOW DO I KNOW IF THIS PROGRAM IS RIGHT FOR ME?
This program assumes that you have had several hours of Python programming experience. Other than that, the only requirement is that you have a curiosity about machine learning. Do you want to learn more about recommendation systems or voice assistants and how they work? If so, then this program is right for you.

ENROLLMENT AND ADMISSION

DO I NEED TO APPLY? WHAT ARE THE ADMISSION CRITERIA?
No. This Nanodegree program accepts all applicants regardless of experience and specific background.

WHAT ARE THE PREREQUISITES FOR ENROLLMENT?
It is recommended that you have the following knowledge, prior to entering the program:

Intermediate Python programming knowledge, including:

• At least 40hrs of programming experience
• Familiarity with data structures like dictionaries and lists
• Experience with libraries like NumPy and pandas is a plus

Basic knowledge of probability and statistics, including:

• Experience calculating the probability of an event
• Knowing how to calculate the mean and variance of a probability distribution is a plus

FAQs Continued

IF I DO NOT MEET THE REQUIREMENTS TO ENROLL, WHAT SHOULD I DO?
You can still succeed in this program, even if you do not meet the suggested requirements. There are a few courses that can help prepare you for the program. For example:

- **Intro to Data Science**
- **Intro to Programming Nanodegree program**

DO I HAVE TO TAKE THE INTRO TO MACHINE LEARNING NANODEGREE PROGRAM BEFORE ENROLLING IN THE MACHINE LEARNING ENGINEER NANODEGREE PROGRAM?
No. Each program is independent of the other. If you are interested in machine learning, you should look at the prerequisites for each program to help you decide where you should start your journey to becoming a machine learning engineer.

TUITION AND TERM OF PROGRAM

HOW IS THIS NANODEGREE PROGRAM STRUCTURED?
The Intro to Machine Learning Nanodegree program is comprised of content and curriculum to support three (3) projects. We estimate that students can complete the program in three (3) months, working 10 hours per week.

Each project will be reviewed by the Udacity reviewer network. Feedback will be provided and if you do not pass the project, you will be asked to resubmit the project until it passes.

HOW LONG IS THIS NANODEGREE PROGRAM?
Access to this Nanodegree program runs for the length of time specified in the payment card above. If you do not graduate within that time period, you will continue learning with month to month payments. See the **Terms of Use** and **FAQs** for other policies regarding the terms of access to our Nanodegree programs.

I HAVE GRADUATED FROM THE BUSINESS ANALYTICS NANODEGREE PROGRAM AND I WANT TO KEEP LEARNING. WHERE SHOULD I GO FROM HERE?
Many of our graduates continue on to our **Machine Learning Engineer Nanodegree** program, and after that, to the **Self-Driving Car Engineer** and **Artificial Intelligence Nanodegree** programs.

FAQs Continued

SOFTWARE AND HARDWARE

WHAT SOFTWARE AND VERSIONS WILL I NEED IN THIS PROGRAM?
You will need a computer running a 64-bit operating system with at least 8GB of RAM, along with administrator account permissions sufficient to install programs including Anaconda with Python 3.x and supporting packages.

Most modern Windows, OS X, and Linux laptops or desktop will work well; we do not recommend a tablet since they typically have less computing power. We will provide you with instructions to install the required software packages.