



COLORADO STATE UNIVERSITY
— GLOBAL —

CSC525: PRINCIPLES OF MACHINE LEARNING

Credit Hours: 3

Contact Hours: This is a 3-credit course, offered in accelerated format. This means that 16 weeks of material is covered in 8 weeks. The exact number of hours per week that you can expect to spend on each course will vary based upon the weekly coursework, as well as your study style and preferences. You should plan to spend 14-20 hours per week in each course reading material, interacting on the discussion boards, writing papers, completing projects, and doing research.

Faculty Information: Faculty contact information and office hours can be found on the faculty profile page.

COURSE DESCRIPTION AND OUTCOMES

Course Description:

In this course, students will examine concepts and theories in machine learning. Students will explore foundational topics that include: supervised and unsupervised learning, learning theory, reinforcement learning, and adaptive control. Students will gain an understanding of applications of machine learning in areas of data mining, human-computer interaction, natural language processing, and computer vision.

Course Overview:

This course will introduce you to Machine Learning (ML) and its applications, and get you started with the basics of developing your own ML-based solutions to a variety of problems. You will learn what types of problems ML can solve and how to categorize them. Additionally, you will receive an introduction to techniques that will help address these different problems. Most importantly, you will have opportunities to practice with the tools and mathematical concepts, which you will need to start programming your own ML applications.

Course Learning Outcomes:

1. Identify applications of machine learning for various areas.
2. Select appropriate machine learning techniques for a given scenario.
3. Identify the characteristics of a given dataset for solving a machine learning problem.
4. Implement machine learning techniques for a given application.
5. Apply learning theory techniques for a given scenario.
6. Compare strategies to solve an adaptive control problem for a given scenario.
7. Implement a solution to solve an unsupervised learning problem.

PARTICIPATION & ATTENDANCE

Prompt and consistent attendance in your online courses is essential for your success at CSU-Global Campus. Failure to verify your attendance within the first 7 days of this course may result in your withdrawal. If for some reason you would like to drop a course, please contact your advisor.

Online classes have deadlines, assignments, and participation requirements just like on-campus classes. Budget your time carefully and keep an open line of communication with your instructor. If you are having technical problems, problems with your assignments, or other problems that are impeding your progress, let your instructor know as soon as possible.

COURSE MATERIALS

Required:

This course utilizes:

1. Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG. eISBN: 9780134845647 or ISBN: 9780134845623.
2. Web-Based Tools:
 - Python version 3.7 software
 - Pycharm IDE
 - Jupyter
 - Tensorflow 2.0 ML Framework or PyTorch 1.4 ML Framework

NOTE: *All non-textbook required readings and materials necessary to complete assignments, discussions, and/or supplemental or required exercises are provided within the course itself. Please read through each course module carefully.*

COURSE SCHEDULE

Due Dates

The Academic Week at CSU-Global begins on Monday and ends the following Sunday.

- **Discussion Boards:** The original post must be completed by Thursday at 11:59 p.m. MT and Peer Responses posted by Sunday 11:59 p.m. MT. Late posts may not be awarded points.
- **Opening Exercises:** Take the opening exercise before reading each week's content to see which areas you will need to focus on. You may take these exercises as many times as you need. The opening exercises will not affect your final grade.
- **Mastery Exercises:** Students may access and retake mastery exercises through the last day of class until they achieve the scores they desire.
- **Critical Thinking:** Assignments are due Sunday at 11:59 p.m. MT.

WEEKLY READING AND ASSIGNMENT DETAILS

Module 1: Current and Future Applications of Machine Learning

Readings

- Chapters 1-4 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Allen, R. (2019). Five Lessons for Applying Machine Learning. *Research Technology Management*, 62(3), 38-44. <https://doi.org/10.1080/08956308.2019.1587330>.
- Wong, W. (2019). Machine Learning: The Magic is How it Works. *Electronic Design*, 67(4), 12-18.

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: AI-Designed Drugs, Materials, and Products

Choose one of the following two assignments to complete this week. Do not complete both assignments. Identify your assignment choice in the title of your submission.

1. After reading the Required Readings, read the article listed below about the pharmaceutical company Exscientia and their use of AI and machine learning systems to design a new drug that has now entered human testing.
 - Murgia, M. (2020, Jan 29). AI-designed drug to enter human clinical trial for first time. Retrieved from <https://www.ft.com/content/fe55190e-42bf-11ea-a43a-c4b328d9061c>
2. As machine learning techniques become more advanced, we will rely on AI technologies more and more to accomplish the things that we previously thought were impossible, and accomplish the difficult much faster. As AI is used more and more in the design of materials, products, and even medicines, it will become increasingly more tightly woven into our everyday lives and give us many advantages. Address the following questions in essay form.
 - a. Do you see the use of AI in the design of products intended for human consumption as a positive or negative, and why?
 - b. If with the use of AI we can design drugs faster and more reliably than we could before, how might this impact human health on a societal and individual level?
 - c. Do you think the relevance of specialized human expertise in the research and design of new drugs and materials will diminish or remain relevant in the future, and why?

Your paper should be 4-5 pages in length and conform to APA style. The CSU Global Writing Center offers resources on how to format your assignment and cite sources in APA style.

Your submission should also include 4-6 references. The *CSU Global Library* is a good place to find these references.

OPTION #2: Natural Language Processing Transformers

Read the following article:

Suryavansh, M. (2020, Jan 4). 2019 — Year of BERT and Transformer: A short post about BERT and NLP projects in 2019.

Then, take a look at and try your hands at the following online language transformers:

- Talk to Transfer

- GROVER - A State-of-the-Art Defense against Neural Fake News

As natural language processing (NLP) advances, the applications of this new technology will become more varied and profound. Based on the article and your own research into natural language processing, address the questions that follow in essay form.

Questions:

- What are possible future applications of NLP?
- How might NLP affect other areas of the AI field?
- What kinds of ethical and social problems might a convincing NLP agent cause?
- What machine learning techniques are used in NLP and what techniques/technologies have promise for the future? Be as specific as possible.

Your paper should be 4-5 pages in length and conform to *CSU Global Guide to Writing and APA*. Include at least four references in addition to the course textbook. The CSU Global Library is a good place to find these references.

Portfolio Reminder

You will select one of two options for your Portfolio Project: 1) Build your own chatbot utilizing Natural Language Processing or 2) Develop your own Unity project utilizing Unity ML-Agents.

Please read through the instructions for the portfolio assignment and begin research on these topics, including the tools and software you will require.

Module 2: Classification Problems

Readings

- Chapter 6 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Chapter 8 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: KNN Classifier with Iris Data

KNN cluster classification works by finding the distances between a query and all examples in its data. The specified number of examples (K) closest to the query are selected. The classifier then votes for the most frequent label found.

There are several advantages of KNN classification, one of them being simple implementation. Search space is robust as classes do not need to be linearly separable. It can also be updated online easily as new instances with known classes are presented.

A KNN model can be implemented using the following steps:

1. Load the data;
2. Initialise the value of k;
3. For getting the predicted class, iterate from 1 to total number of training data points;
4. Calculate the distance between test data and each row of training data;

5. Sort the calculated distances in ascending order based on distance values;
6. Get top k rows from the sorted array;
7. Get the most frequent class of these rows; and
8. Return the predicted class.

For your assignment, you will build a KNN classifier in Python.

Download the class data in CSV format.

Your classifier should be able to, using this data, predict the type of iris based on the sepal length and width (the parts of the calyx) and the petal length and width, in centimeters.

Submission should include an executable Python file that accepts input of 4 floating point numbers representing, respectively, sepal length, sepal width, petal length, and petal width.

OPTION #2: Predicting Video Game Preferences with KNN

KNN cluster classification works by finding the distances between a query and all examples in its data. The specified number of examples (K) closest to the query are selected. The classifier then votes for the most frequent label found.

There are several advantages of KNN classification, one of them being simple implementation. Search space is robust as classes do not need to be linearly separable. It can also be updated online easily as new instances with known classes are presented.

A KNN model can be implemented using the following steps:

1. Load the data;
2. Initialise the value of k;
3. For getting the predicted class, iterate from 1 to total number of training data points;
4. Calculate the distance between test data and each row of training data;
5. Sort the calculated distances in ascending order based on distance values;
6. Get top k rows from the sorted array;
7. Get the most frequent class of these rows; and
8. Return the predicted class.

For your assignment, you will build a KNN classifier in Python.

Download the class data in CSV format:

Using this data, your classifier should be able to predict a person's favorite video game genre based on their age, height, weight, and gender. (Do not worry about real-world accuracy here. This is to provide you an opportunity to practice.)

You can also choose to collect and use your own data points.

Submission should include an executable Python file that accepts input of 4 floating point numbers representing, respectively, age (in years), height (in inches), weight (in lbs), and gender (females represented by 0s and males represented by 1s).

Portfolio Milestone (25 points)

OPTION #1: Beginning Research on Unity ML-Agents

It is time to begin research for your portfolio project.

Unity is a free to use (on a personal basis) development platform often used for video games or, as in our case, an experimental simulation platform.

Unity can be downloaded free of charge at [Unity.com](https://unity.com)

If you are new to Unity, there are abundant resources for beginners. Please see the following resources to get started:

- Unity Documentation
- Unity Learn

Unity ML-Agents is a deep learning software library that provides an application interface between Unity and Tensorflow neural network inference and training in Python.

Unity ML Agents can be downloaded on Github from the following repository:

[Unity-Technologies/ml-agents](https://github.com/Unity-Technologies/ml-agents): Unity Machine Learning Agents Toolkit

In addition to resources found directly on the above linked Github repository, the following resources may also come in handy:

- Learn how to use Unity Machine Learning Agents
- Unity Machine Learning
- An Introduction to Unity ML-Agents

For your portfolio milestone, please install and explore Unity and Unity ML Agents.

Then write no less than a page describing your experience running through one of Unity ML Agents sample scenarios. Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

OPTION #2: Beginning Research on Natural Language Processing and Chatbot Technology

It is time to begin research for your portfolio project.

Chatbots use in customer service is currently growing at an accelerating rate. Especially when used within specific domains, chatbots can free up human beings for many customer-service and information-retrieval oriented tasks. More and more, they will be used to control our home appliances, schedule our calendars, and even entertain and socialize with us.

For your portfolio project, you are going to build a chatbot using Natural Language Processing principles. Your chatbot does not have to utilize speech to text (but can!), but will respond to text input with a text output.

Your chatbot will be a learning-based chatbot, as opposed to a rule-based chatbot. There are two types of learning-based chatbots:

1. Retrieval-based chatbots. These chatbots learn to select which response to which user queries. These function similarly to rule-based chatbots, but use machine learning techniques in the process of selecting a response instead of using predefined rules.
2. Generative chatbots. These chatbots generate each response on the fly. Currently, these are more likely to be used for entertainment, but the market for more advanced generative chatbots grows as NLP technologies advance.

The following tools and resources should be researched to plan your NLP chatbot. The libraries below are compatible with one another.

Libraries, APIs and Python Tools:

- NLTK: Natural Language Toolkit — NLTK 3.4.5 documentation
- Chatterbot: ChatterBot Tutorial
- TextBlob: TextBlob: Simplified Text Processing — TextBlob 0.15.2 documentation
- spaCy: spaCy · Industrial-strength Natural Language Processing in Python

Most learning chatbots are trained using transformers, like seq2seq encoder-decoders. The practice of using transformers for NLP began in 2017 with Vaswani et al's (2017) *Attention is all you Need*. Please click on "PDF" in the menu on the right to download the file.

Transformer libraries can be found collected at the Github repository.

For your portfolio milestone, write no less than a page describing a general plan for your chatbot, such as what tools and libraries you will use, whether it will be an open-domain or closed-domain bot, and whether it will be a retrieval-based bot or a generative bot. Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

Module 3: Linear and Nonlinear Regression

Readings

- Chapter 7 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Chapter 9 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Terhanian, G. (2019). The possible benefits of reporting percentage point effects. *International Journal of Market Research*, 61(6), 635–650. <https://doi.org/10.1177/1470785319838742>. Discussion (25 points)

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: Simple Linear Regression in Scikit Learn

Linear regression is a supervised learning algorithm that predicts a dependent variable value based on an independent variable by fitting a linear equation to the data.

There are several advantages to linear regression, mainly high efficiency. This efficiency can easily lead to overfitting the data, however.

For your assignment, you will build a linear regression model in Python.

The Boston housing dataset can be loaded in scikit-learn using the command `load_boston()` after `from sklearn.datasets import load_boston`.

Using this data, our model should be able to predict the value of a house using the features given in the dataset.

The Boston housing dataset is a common target for regression analysis, feel free to use Google to conduct your own research for this assignment. Feel free to also use the following documentation and resources:

- Linear regression
- Linear Regression on Boston Housing Dataset

Submission should include an executable Python file demonstrating the prediction of housing prices.

OPTION #2: Simple Polynomial Regression in Python

Polynomial regression is a form of nonlinear regression that describes nonlinear relationships in a dataset.

There are several advantages to linear regression, mainly high accuracy.

For your assignment, you will build a polynomial regression model in Python.

Please download the Positions Salaries dataset in CSV format:

Using this data, our model should be able to predict the value of an employee candidate given their years of experience.

Consider using Google to conduct your own research for this assignment. Feel free to also use the following documentation and resources:

- Robust nonlinear regression in scipy
- Machine Learning: Polynomial Regression with Python

Submission should include an executable Python file demonstrating the prediction of employee salary based on years of experience.

Portfolio Milestone (25 points)

OPTION #1: ML-Agents Project Draft

Write at least a page describing, in detail, your plan for your final portfolio project. Describe how your demo will demonstrate an adaptive control solution in an uncertain environment. Describe in as much detail as possible the goal you will be training your agents to achieve.

Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

OPTION #2: NLP Chatbot Project Draft

Write at least a page describing, in detail, your plan for your final portfolio project. Describe how your chatbot will demonstrate an adaptive control solution in an uncertain environment. Describe, in as much detail as possible, the goals of your chatbot, the tools, libraries, and models you intend to use for your chatbot, and the general structure of your program.

Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

Module 4: Working with Neural Networks

Readings

- Chapter 15 Section 5 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.

- Al-Bdour, G., Al-Qurran, R., Al-Ayyoub, M., & Shatnawi, A. (2019). A detailed comparative study of open source deep learning frameworks. Retrieved from <https://arxiv.org/abs/1903.00102?>
- Gerasimovic, M., & Bugarcic, U. (2018). Enrollment Management Model: Artificial Neural Networks versus Logistic Regression. *Applied Artificial Intelligence*, 32(2), 153–164. <https://doi.org/10.1080/08839514.2018.1448146>

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: Build a Tensorflow Demo

Please see the following Tensorflow examples:

- TensorFlow-Examples/tensorflow_v2 at master · aymericdamien/TensorFlow-Examples

For your assignment, choose one of the demonstrations provided in the link above.

The purpose of this assignment is to gain hands-on experience running a Tensorflow model and demonstrate the ability to use the tools and documentation available.

1. Before continuing, please see the following documentation and install Jupyter. Installing Jupyter Notebook
2. Then, make sure you have Tensorflow installed if it is not already. We recommend version 2.1. We also recommend selecting the CPU version if you are working on a notebook, or the GPU version only if you have an NVIDIA GPU with CUDA compute capability of at least 5.0. Also note that, if installing the GPU version, CUDA drivers must also be installed. This is an extra step that may provide huge performance improvements at the risk of a more complicated setup process. Install TensorFlow with pip
3. Download the raw .ipynb file for the example you chose. (Hint: On the Github page of the example you chose, right click "Raw", then "Save As" to save the file)

The screenshot shows a GitHub repository for 'aymericdamien / TensorFlow-Examples'. The file 'recurrent_network.ipynb' is selected. A context menu is open over the 'Raw' link, with 'Save link as...' highlighted. Below the file name, there is a commit message 'aymericdamien add TensorFlow v2 RNN examples' and a contributor list. The main content area shows the beginning of the notebook, titled 'Recurrent Neural Network Example', with an 'RNN Overview' section. The overview includes a diagram of an unrolled RNN cell and a sequence of four cells labeled h_0 , h_1 , h_2 , and h_3 .

4. Open the .ipynb file using Jupyter Notebook.
5. Run through the example cells one-by-one until you have reached the end result.
6. Submit an essay addressing the following questions:
 - What was notable during the installation of Tensorflow? Did you run into any problems?

- Describe your chosen example. How did the example run? What did it do?
- What type of model does your example use? Do further research to describe the model type and its common applications.
- How might you leverage techniques found in your example for your portfolio project?
- What type of dataset did the example use? Is it possible to improve that dataset? What other ways might the model type used in the example be expanded, improved upon, or integrated into other techniques?
- Is the model used an example of a supervised or unsupervised model?

Your paper should be 4-6 pages in length and conform to *CSU Global Guide to Writing and APA*. Include at least four references. The CSU Global Library is a good place to find these references.

OPTION #2: Build a Pytorch Demo

Please see the following examples and tutorials:

- [pytorch/examples](#): A set of examples around pytorch in Vision, Text, Reinforcement Learning, etc.
- [Welcome to PyTorch Tutorials — PyTorch Tutorials 1.4.0 documentation](#)

For your assignment, choose one of the demonstrations provided in the link above.

The purpose of this assignment is to gain hands-on experience running a Tensorflow model and demonstrate the ability to use the tools and documentation available.

1. Using your chosen package manager such as pip, make sure you have PyTorch installed if it is not already. We recommend version 1.4.
2. Download the scripts from your chosen example as found on the example's Github page.
3. Run the scripts using the instruction on the example's Github page.
4. Submit an essay addressing the following questions:
 - What was notable during the installation of PyTorch? Did you run into any problems?
 - Describe your chosen example. How did the example run? What did it do?
 - What type of model does your example use? Do further research to describe the model type and its common applications.
 - How might you leverage techniques found in your example for your portfolio project?
 - What type of dataset did the example use? Is it possible to improve that dataset? What other ways might the model type used in the example be expanded, improved upon, or integrated into other techniques?
 - Is the model used an example of a supervised or unsupervised model?

Your paper should be 3-5 pages in length and conform to *CSU Global Guide to Writing and APA*. Include at least four scholarly references. The CSU Global Library is a good place to find these references.

Portfolio Milestone (25 points)

OPTION #1: ML Agents Project Setup

It is time to set up the training environment for your agent training. Please refer to the following documentation in the ML Agents Github repository for guidance:

<https://github.com/Unity-Technologies/ml-agents/blob/master/docs/Readme.md>

Write at least a page describing, in detail, your training environment. Discuss how you expect your agent to learn from its environment. Discuss the hyperparameters you have decided to test for your first training. Will your agents use LSTM/RNN memory? How will you reward your agents to motivate them to achieve their goal? Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

OPTION #2: NLP Chatbot Project Data Collection

It is time to collect conversation data for your NLP chatbot. The type of dataset you choose will depend on the libraries and model you have chosen to utilize. Below are some useful choices of NLP training datasets:

- Question-Answer Dataset
- Microsoft Research WikiQA Corpus
- Yahoo Language Data
- TREC QA Collection
- Ubuntu Dialogue Corpus
- Relational Strategies in Customer Service Dataset
- Customer Support Tweets on Twitter
- Semantic Web Interest Group IRC Chat Logs
- Cornell Movie Dialogue Corpus
- Conversational Intelligence Challenge 2 Data
- Santa Barbara Corpus of Spoken American English
- The NPS Chat Corpus
- Microsoft Research Goal Oriented Frame Tracking Dataset
- MultiWOZ Corpus

Write at least a page describing the dataset(s) you chose, and describe in detail how you intend to train your model on the dataset and how your chatbot will utilize that training. Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

Module 5: Evaluation and Dataset Optimization: Overfitting, Underfitting, Variance, and Bias

Readings

- Chapter 5 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Chaudhary, N., & Chowdhury, D. R. (2019). Data Preprocessing for Evaluation of Recommendation Models in E-Commerce. <https://doi.org/10.3390/data4010023>
- Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2013). The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks. *Procedia Technology*, 11, 32–39. <https://doi.org/10.1016/j.protcy.2013.12.159>
- Neal, B. (2019). On the Bias-Variance Tradeoff: Textbooks Need an Update. Retrieved from [https://arxiv.org/abs/1912.08286?](https://arxiv.org/abs/1912.08286)

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: Text Dataset Augmentation

State-of-the-art machine learning models use millions, sometimes billions, of parameters. When training our models, the number of distinct examples we need is proportional to the number of parameters our model has.

Invariance is a trait of a neural network model (for instance, an image classification model) that can robustly classify objects even when the objects are placed in different orientations, have different sizes, and illumination differences. Many of these potential differences in our dataset can be created artificially, instead of collecting more images. Methods could be to alter brightness or contrast of the image, stretch or skew operations, or a variety of translation methods.

But what about textual data for NLP projects? Possibilities include synonym replacement, which replaces words with words that mean the same thing, random inserts, swaps, and deletions, among others.

For your assignment, submit a Python script that will take any text dataset and augment it in some way to expand the dataset. Submission must include a script that will augment any text dataset within it's folder. Please include a sample un-augmented dataset with your submission.

OPTION #2: Image Dataset Augmentation

State-of-the-art machine learning models use millions, sometimes billions, of parameters. When training our models, the number of distinct examples we need is proportional to the number of parameters our model has.

Invariance is a trait of a neural network model (for instance, an image classification model) that can robustly classify objects even when the objects are placed in different orientations, have different sizes, and illumination differences. Many of these potential differences in our dataset can be created artificially, instead of collecting more images. Methods could be to alter brightness or contrast of the image, stretch or skew operations, or a variety of translation methods.

For your assignment, submit a Python script that will take any image dataset and augment it in some way to expand the dataset. Submission must include a script that will augment any image files within it's folder. Please include a small sample un-augmented dataset with your submission.

Portfolio Milestone (25 points)

OPTION #1: ML Agents Project Training

Time to train your agents!

Your training scenario should be ready, as should your reward functions (as explained in the ML Agents documentation).

Your milestone submission will be the Tensorflow graphs from your project (which you will have learned about using in the ML Agents documentation). In order to download the graphs, just check the "Data download links" option on the upper-left in TensorBoard, and then click on the "CSV" button that will appear under your scalar summary. Submit the graphs in CSV format in a zip file.

OPTION #2: NLP Chatbot Training

Time to train your chatbot!

How your training will proceed will depend on the tools you are using, the model type you are training, and how you have set up your model within your program. You have your text dataset, and should have done enough research. You are ready to train your NLP model.

For your milestone, write at least a page describing your experience training your model. Describe in detail the type of model you are using and the dataset you are training it on. Discuss the hyperparameters used, and how your first training worked out.

Module 6: Hyperparameter Tuning and Feature Engineering

Readings

- Chapters 10-11 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.

- Chapter 13 Sections 1 and 2 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Chapter 14 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Chang, D. T. (2019). Bayesian Hyperparameter Optimization with BoTorch, GPyTorch and Ax. Retrieved from [https://arxiv.org/abs/1912.05686?](https://arxiv.org/abs/1912.05686)

Discussion (25 points)

Critical Thinking (75 points)

OPTION #1: Essay - Feature Engineering and Hyperparameter Tuning

Hyperparameters are variables that govern the training of the machine learning process. Finding the optimal set of hyperparameters is essential to producing a model that optimally reduces the loss function. Hyperparameter tuning or optimization is the process of selecting an optimal set of hyperparameters, often using machine learning techniques itself.

Several methods are often used for hyperparameter optimization methods. One of the least efficient is grid search. Other common methods include Random Search, in which provides a statistical distribution for each hyperparameter from which values may be randomly sampled, and Bayesian optimization, which uses each iteration to improve the hyperparameters of the next. Other methods include using SVMs, logistic regression, or neural networks specifically for choosing optimal hyperparameters, and evolutionary optimization, which uses evolutionary algorithms to converge on a set of hyperparameters.

"A Gaussian process analysis of the function from hyper-parameters to validation set performance reveals that for most data sets only a few of the hyper-parameters really matter, but that different hyper-parameters are important on different data sets. This phenomenon makes grid search a poor choice for configuring algorithms for new data sets." - Bergstra, 2012

A feature is an attribute shared by all independent units on which analysis or prediction is being done. Any attribute can be a feature, as long as it is useful to the model. Feature engineering is the process of using domain-specific knowledge to extract features from raw data contextually, using a variety of data-mining techniques. For example, in order for a recommender system to be trained properly, it needs to know what domain-specific knowledge about its users preferences or needs will lead it to a viable prediction. Another example would be vision systems. If the system is built to detect a certain type of object, explicitly programming detection of features specific to that object can greatly enhance accuracy. Feature engineering is highly dependent on the problem at hand.

For your assignment, choose one of the following hypothetical use cases.

- Text classification model intended to identify fake news and/or abusive content;
- Decision Support System (DSS) for corporate decision making at a financial firm;
- AI-based recruitment enhancement and human resources management tools;
- NLP-based Chatbot intended for psychological therapy and diagnosis;
- Image classification model for classifying and diagnosing malignant melanoma and differentiating from benign moles; or
- Movie recommender system for a popular streaming service.

Write an essay addressing, on purely theoretical grounds, the following questions in regards to your chosen hypothetical use-case:

- What methods of hyperparameter tuning might your example benefit from? In order to answer this question, the following questions must also be addressed:

- What type of machine learning model might your example use? What methods will it use to make its predictions or classification decisions?
- What specific hyperparameters might apply to the models used?
- What ways might feature engineering be used to enhance the example hypothetical models accuracy and efficiency? Based on the specific problem, what features might be extracted in order to boost the models predictions?

Remember, you are examining these hypothetical examples from a theoretical standpoint. You do not need to base your answers on actual implementations of these exact types of systems. However, you will be required to provide references to the ideas and concepts used in your research on hyperparameter optimization and feature engineering and extraction.

Your paper should be 4-6 pages in length and conform to *CSU Global Guide to Writing and APA*. Include at least four references. The CSU Global Library is a good place to find these references.

OPTION #2: Presentation - Feature Engineering and Hyperparameter Tuning

Hyperparameters are variables that govern the training of the machine learning process. Finding the optimal set of hyperparameters is essential to producing a model that optimally reduces the loss function. Hyperparameter tuning or optimization is the process of selecting an optimal set of hyperparameters, often using machine learning techniques itself.

Several methods are often used for hyperparameter optimization methods. One of the least efficient is grid search. Other common methods include Random Search, in which provides a statistical distribution for each hyperparameter from which values may be randomly sampled, and Bayesian optimization, which uses each iteration to improve the hyperparameters of the next. Other methods include using SVMs, logistic regression, or neural networks specifically for choosing optimal hyperparameters, and evolutionary optimization, which uses evolutionary algorithms to converge on a set of hyperparameters.

"A Gaussian process analysis of the function from hyper-parameters to validation set performance reveals that for most data sets only a few of the hyper-parameters really matter, but that different hyper-parameters are important on different data sets. This phenomenon makes grid search a poor choice for configuring algorithms for new data sets." - Bergstra, 2012

A feature is an attribute shared by all independent units on which analysis or prediction is being done. Any attribute can be a feature, as long as it is useful to the model. Feature engineering is the process of using domain-specific knowledge to extract features from raw data contextually, using a variety of data-mining techniques. For example, in order for a recommender system to be trained properly, it needs to know what domain-specific knowledge about its users preferences or needs will lead it to a viable prediction. Another example would be vision systems. If the system is built to detect a certain type of object, explicitly programming detection of features specific to that object can greatly enhance accuracy. Feature engineering is highly dependent on the problem at hand.

For your assignment, choose one of the following hypothetical use cases:

- Text classification model intended to identify fake news and/or abusive content;
- Decision Support System (DSS) for corporate decision making at a financial firm;
- AI-based recruitment enhancement and human resources management tools;
- NLP-based Chatbot intended for psychological therapy and diagnosis;
- Image classification model for classifying and diagnosing malignant melanoma and differentiating from benign moles; or
- Movie recommender system for a popular streaming service.

Create a presentation addressing, on purely theoretical grounds, the following questions in regards to your chosen hypothetical use-case:

- What methods of hyperparameter tuning might your example benefit from? In order to answer this question, the following questions must also be addressed:
 - What type of machine learning model might your example use? What methods will it use to make its predictions or classification decisions?
 - What specific hyperparameters might apply to the models used?
- What ways might feature engineering be used to enhance the example hypothetical models accuracy and efficiency? Based on the specific problem, what features might be extracted in order to boost the models predictions?

Remember, you are examining these hypothetical examples from a theoretical standpoint. You do not need to base your answers on actual implementations of these exact types of systems. However, you will be required to provide references to the ideas and concepts used in your research on hyperparameter optimization and feature engineering and extraction.

Your PowerPoint should consist of a minimum of 6-8 slides. Support your analysis using at least four authoritative sources, cited using correct APA styling as found in *CSU Global Guide to Writing and APA*.

Submit your PowerPoint presentation using the file name

CSC525_Module6CriticalThinking_Option2_<LastName>_<FirstName>.pptx

Portfolio Milestone (25 points)

OPTION #1: ML Agents Project Alpha

You should have a working alpha version of your ML Agents project by now.

For your milestone, submit at least a page comparing and contrasting what your project does now with what you would like it to do in your final submission. Discuss what it does well and what it could do better. Speculate on how you might improve it. Will it require retraining? Can you leverage any other techniques to improve it? Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

OPTION #2: NLP Chatbot Project Alpha

You should have a working alpha version of your NLP chatbot project by now.

For your milestone, submit at least a page comparing and contrasting what your project does now with what you would like it to do in your final submission. Discuss what it does well and what it could do better. Speculate on how you might improve it. Will it require retraining? Can you leverage any other techniques to improve it? Your submission should be at least a page in length and conform to *CSU Global Guide to Writing and APA*.

Module 7: Unsupervised and Self-Supervised Learning: From DBNs and Autoencoders to K-Means Clustering

Readings

- Chapter 13 Section 3 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Retrieved from <https://arxiv.org/abs/1706.03762>

Discussion (25 points)

Module 8: The Future: Multimodal Models, Voting Ensembles, and Federated Learning

Readings

- Chapter 12 in Fenner, M. (2020). *Machine learning with Python for everyone*. Boston, MA: Addison-Wesley Professional PTG.
- Khargonekar, P. P., & Dahleh, M. A. (2018). Advancing systems and control research in the era of ML and AI. *Annual Reviews in Control*, 45, 1–4. <https://doi.org/10.1016/j.arcontrol.2018.04.001>
- Lin Li, & Chun-Kit Ngan. (2017). A weight-adjusted-voting framework on an ensemble of classifiers for improving sensitivity. *Intelligent Data Analysis*, 21(6), 1339–1350. <https://doi.org/10.3233/IDA-163184>

Discussion (25 points)

Portfolio Project (225 points)

OPTION #1: ML Agents Project Final Version

Please submit your final ML Agents project in the form of either an executable file demonstrating the scenario, or a video or link to a video demonstration of your ML Agents scenario. Your agents must meet the following conditions:

- Your agents must estimate outputs necessary to meet the stated goal.
- They must deal with an alterable, changeable environment of some kind. They shouldn't be trained or operate in a static environment.

In addition to your program, your submission should include a 4-6 page essay describing the final version of your demonstration and the methods you utilized to compel your agent to achieve its defined goal in the environment you set up and how you set up the reward signal to guide the agents behavior. In your paper, please address the following details:

- Problems you encountered structuring the reward signal.
- Did you utilize any form of transfer learning?
- Did you use any form of voting ensemble or federated learning to accomplish the goal? Does your agent utilize any form of meta-learning or self-tuning?
- How did you decide on hyperparameters?

Important: Your final submission will also be required to include at least 12 sources from which your work is based. Please feel free to utilize the CSU Global Library to find resources to help you. Your submission should conform to *CSU Global Guide to Writing and APA*.

Please include with your submission at least a paragraph stating the goal the agents should be achieving as well as any instructions for running the simulation.

OPTION #2: NLP Chatbot Final Version

Please submit your final NLP chatbot project in the form of an executable file or link allowing the instructor to chat with your chatbot. Your chatbot must meet the following conditions:

- Use NLP learning methods to respond to user inputs.
- Chatbot responses should not be nonsense. Responses should be clearly in response to the input text.

Please include with your submission at least a paragraph stating whether the chatbot is open-domain or closed, what tools and libraries used, and what type of NLP model is used in the chatbot, as well as any instructions for running or accessing the chatbot.

COURSE POLICIES

Grading Scale	
A	95.0 – 100
A-	90.0 – 94.9
B+	86.7 – 89.9
B	83.3 – 86.6
B-	80.0 – 83.2
C+	75.0 – 79.9
C	70.0 – 74.9
D	60.0 – 69.9
F	59.9 or below

Course Grading

20% Discussion Participation
45% Critical Thinking Assignments
35% Final Portfolio Project
0% Live Classroom

IN-CLASSROOM POLICIES

For information on late work and incomplete grade policies, please refer to our [In-Classroom Student Policies and Guidelines](#) or the Academic Catalog for comprehensive documentation of CSU-Global institutional policies.

Academic Integrity

Students must assume responsibility for maintaining honesty in all work submitted for credit and in any other work designated by the instructor of the course. Academic dishonesty includes cheating, fabrication, facilitating academic dishonesty, plagiarism, reusing /re-purposing your own work (see *CSU-Global Guide to Writing and APA Requirements* for percentage of repurposed work that can be used in an assignment), unauthorized possession of academic materials, and unauthorized collaboration. The CSU-Global Library provides information on how students can avoid plagiarism by understanding what it is and how to use the Library and Internet resources.

Citing Sources with APA Style

All students are expected to follow the *CSU-Global Guide to Writing and APA Requirements* when citing in APA (based on the APA Style Manual, 6th edition) for all assignments. For details on CSU-Global APA style, please review the APA resources within the CSU-Global Library under the “APA Guide & Resources” link. A link to this document should also be provided within most assignment descriptions in your course.

Disability Services Statement

CSU-Global is committed to providing reasonable accommodations for all persons with disabilities. Any student with a documented disability requesting academic accommodations should contact the Disability Resource Coordinator at 720-279-0650 and/or email ada@CSUGlobal.edu for additional information to coordinate reasonable accommodations for students with documented disabilities.

Netiquette

Respect the diversity of opinions among the instructor and classmates and engage with them in a courteous, respectful, and professional manner. All posts and classroom communication must be conducted in accordance with the student code of conduct. Think before you push the Send button. Did you say just what you meant? How will the person on the other end read the words?

Maintain an environment free of harassment, stalking, threats, abuse, insults or humiliation toward the instructor and classmates. This includes, but is not limited to, demeaning written or oral comments of an ethnic, religious, age, disability, sexist (or sexual orientation), or racist nature; and the unwanted sexual advances or intimidations by email, or on discussion boards and other postings within or connected to the online classroom. If you have concerns about something that has been said, please let your instructor know.