

We will use ChatGPT to learn background information about machine learning, types of models, and the process of training models. After this, we will ask ChatGPT to define a neural network model that we can use for a scientific task. The list below is a sample of prompts that could be used for this purpose. Note that the answers given by ChatGPT rely on a random seed and previously asked questions – everyone's answers will likely be different!!

1. I am a materials scientist, and I would like to use machine learning to understand my experimental data. I will ask you questions about how I can use machine learning in my research and I'd like you to respond with bullet points, each containing one or two sentences. To start, can you give an overview of the five most important machine learning concepts that I need to understand?
2. Can you explain the process of training a model?
3. Explain the optimization process in two bullet points
4. Can you list 5 major classes of machine learning models, and provide one advantage and one disadvantage for each?
5. Are any of these types of models compatible with image data?
6. Please explain how convolutional neural networks work in three steps. Then please give a summary of why CNNs are useful for handling images.
7. I'm doing an experiment where aggregates of particles are injected into a chamber which is illuminated by x-rays, and x-ray diffraction data is collected from scattering off the aggregates. I'd like to use machine learning to predict the number of particles in each aggregate from the diffraction patterns. What do I need to consider about data preparation before I start training a model?
8. I'd like to start by generating synthetic data. Write a python script using numpy arrays that generates $N \times N$ images with 1-5 randomly-placed circles of a user-provided radius, all of which either touch or overlap one another. Ensure that the first circle is placed within the center $N/2 \times N/2$ pixels of the image
9. Make a neural network model in PyTorch that classifies $N \times N$ grayscale images into 5 classes
10. Write the training loop to optimize the SimpleClassifier which displays a progress bar and prints the current accuracy on the training and validation datasets for each epoch. Also, save the training and validation losses and accuracies at epoch to lists.