Pipeline of Machine Learning

Chris Cornwell

Aug 26, 2025

Outline

Outline

Steps in an ML Project

Project Pipeline

0. Define the problem.
1. Collect data.
2. Design the features in the data.
3. Training of the model.
4. Test the model.

Discussion on the ML Project Pipeline

0. Define the problem.
    - Example: From a given image, determine if it is an image of a dog or not.
1. Collect data.
    - Example: Put together a large collection of images, some having dogs in them, others having a different animal, or no animal. Have a label (your output, "y") for each image. Split into training set and test set.
2. Design the features in the data.
    - Not one thing that you always do here. Sometimes use experience/knowledge of what the data represents, sometimes use another learning algorithm to learn good features.

Discussion on the ML Project Pipeline

3. Training of the model.
    - Model is determined by set of parameters. In training, you alter the parameters iteratively – “tune” them – using optimization techniques (on the loss function).
4. Test the model.
    - Evaluate the trained model’s performance on test data, measured by the same loss function.

Difficulty in Defining the Problem

While you often have a clearly defined problem when first learning ML (here are images of handwritten numbers; determine which number is written), this is not typical of real settings.

> Example. You have a database of many Tweets (from X / Twitter) about news events and you are interested in using machine learning to determine when a Tweet about news is misinformation versus when it is not.
> What will your predictive model do? Is it a simple classification problem? If so, what are the classes? If not, what alternative is there?

Sometimes the issue is in the data.

> Example. A company wants to try to make a predictive model, using courses taken, types of work experience, other personal information, etc to predict how good a job candi-

Difficulty in Designing Features

Talk about domain knowledge, like it discusses in the textbook. Do some filter examples of images.

Difficulty in Designing Features

Talk about domain knowledge, like it discusses in the textbook. Do some filter examples of images.

Training the model

Where we will spend a lot of time; lots of developed mathematics and algorithms.

Questions?