Machine Learning Course Workbook

Introduction

ML is everywhere!

Where (else) do you use ML in your everyday life incl. work?

-

ML history: Why now?

What accelerated the rise of ML in the last few years?

-
-
-

What is the difference between ANI and AGI?

What is ML?

Define ML:

What do you need to create an ML-powered product (i.e., value)?

-
-
-
-

AI and ML Researchers, Statisticians, and Data Scientists all use a certain set of tools.
What is the difference between...

- ➔ ML vs. AI?
- ➔ ML vs. Deep Learning?
- ➔ ML vs. Statistics?
- ➔ ML vs. Data Science?

How do machines "learn"?

Describe the different learning strategies:

- ➔ Unsupervised Learning:
- ➔ Supervised Learning:
- ➔ Reinforcement Learning:

When should you use ML?

When should you not use ML?

For which kinds of problems does ML have a high chance of success and when is the outcome uncertain?

What distinguishes an ML project from a data science project (in terms of deliverables)?

In what ways can you create value with ML?

Solving problems with ML: Workflow

What are the 3 main steps to create value with ML?

1.
2.
3.

What are the two main deployment possibilities for an ML model and when should you use which?

-
-

Which tasks take up most of a Data Scientist's time?

ML with Python

What are the standard abbreviations used when importing the numpy and pandas libraries?

```
import numpy as ...
import pandas as ...
```

Data & Preprocessing

What are "features" and what are "labels"?

- ➔ Features:
- ➔ Labels:

What does structured and unstructured data look like? Which of them is homogeneous and which (usually) heterogeneous?

- ➔ Structured Data:
- ➔ Unstructured Data:

What is the difference between feature extraction and feature engineering?

- ➔ Feature Extraction:
- ➔ Feature Engineering:

A feature matrix X *has the shape* (n x d). *What do* n *and* d *stand for?*

- ➔ *n:* number of ...
- ➔ *d:*

What constitutes 1 data point?

You are given a dataset with time series data, consisting of measurements from d *sensors for* n *time points. What would your feature matrix look like, if your task is…*

- ➔ … to make a prediction for each time point?
- ➔ … to categorize the different sensors?
- ➔ … to predict the quality of each of the 100 products that were produced during this time span?

Feature Extraction

What is one way to transform categorical features into a meaningful numerical representation?

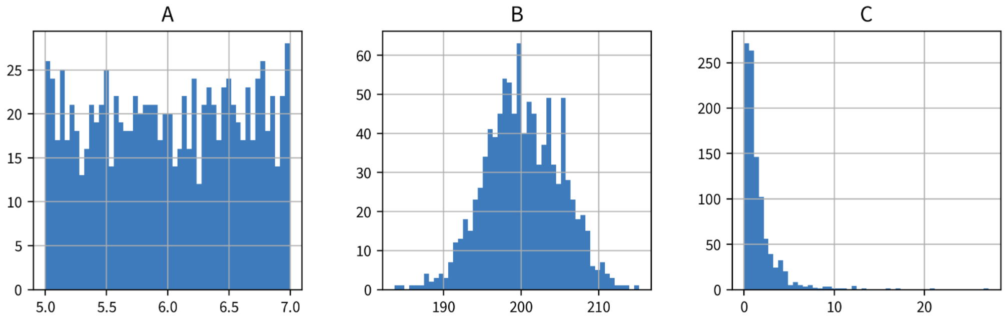
What are the steps to transform a corpus (i.e., dataset with text documents) into a TF-IDF feature matrix?

What are the disadvantages of TF-IDF feature vectors?

-
-

Feature Engineering & Transformations

These are the histograms of 3 different variables A, B, *and* C:



How would you characterize their distributions (Gaussian, exponential, uniform) and which kind of transformation (StandardScaler, MinMaxScaler, PowerTransformer) would be best suited for which of the variables?

➔ A:

➔ B:

➔ C:

Computing Similarities

What preprocessing steps can be helpful to compute a more meaningful similarity or distance between your data points' feature vectors (especially for heterogeneous data)?

-
-

Garbage in, garbage out!

Think about some of the datasets you've encountered in the past: In what ways were they messy?

Which concrete next steps should your organization take to improve their data quality?

ML Solutions: Overview

What does the output of the different algorithm categories look like for one data point?

- Dimensionality Reduction:
- Anomaly Detection:
- Clustering:
- Classification:
- Regression:

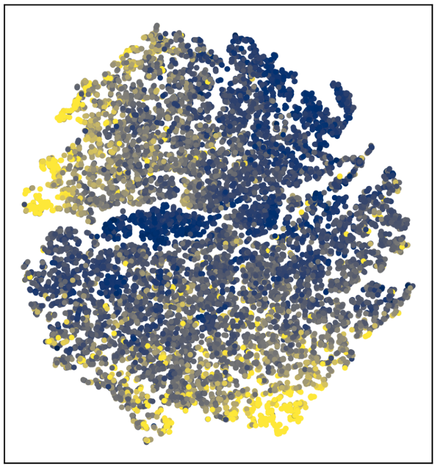
What are the benefits of breaking down a complex input-output problem into simpler subproblems?

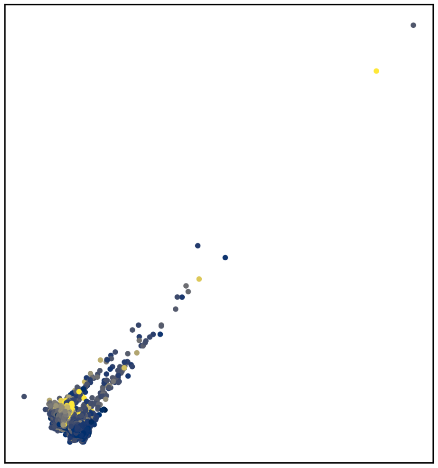
-

Unsupervised Learning

Dimensionality Reduction

Guess: Which plot was generated with PCA and which with t-SNE?





How does PCA work?

Is PCA using the original input features for the computation or does it first compute a similarity matrix for the data points? What about Kernel PCA?

How does t-SNE work?

Is t-SNE using the original input features for the computation or does it first compute a similarity matrix for the data points?

When would you use PCA and when would you use t-SNE?

In the notebook, what did you observe about the PCA eigenvalue spectrum for the data with and without outliers? How do you interpret this?

Outlier/Anomaly Detection

What factors should you consider when choosing an outlier detection method?

How does the γ-index work?

How could you set the parameter k *of the γ-index to detect a cluster of outliers?*

Clustering

How does the k-means algorithm work?

True or False: One disadvantage of k-means is that it assumes spherical clusters?

K-means: What would happen if you set k *to a very large value, e.g., the number of data points?*

How does the DBSCAN algorithm work?

What are the advantages of DBSCAN?

Supervised Learning Basics

Different types of models

What is the difference between a regression and a classification problem?

How can you tell if a classification or regression dataset is linear or non-linear?

When should you use a features-based and when a similarity-based model and what are their respective drawbacks?

Model Evaluation

Name three regression evaluation metrics:

-
-
-

Name two classification evaluation metrics:

-
-

What is the stupid baseline you should always compare your regression models against?

What is the stupid baseline you should always compare your classification models against?

When should you absolutely evaluate your models with the balanced accuracy metric?

How does a cross-validation work? What are the advantages and disadvantages compared to using a fixed validation set?

Supervised Learning Models

Linear Models

How does a linear model compute the prediction for a new data point?

What happens when you use a regularized model and set the regularization parameter to a high value (e.g., alpha for a linear ridge regression model)?

Decision Trees

How does a decision tree compute the prediction for a new data point?

For a decision tree with max_depth=2, how many different features can be used at most for the prediction?

Ensemble Methods

How does a random forest compute the prediction for a new data point?

k-Nearest Neighbors (kNN)

How does a kNN model compute the prediction for a new data point?

Why is it better to use an odd number of nearest neighbors for kNN for a binary classification problem?

Kernel Methods

How does a SVM compute the prediction for a new data point?

Why is it more efficient to compute the prediction for a new data point using a support vector regression (SVR) model compared to a kernel ridge regression model?

Deep Learning & more

Information Retrieval (Similarity Search)

What is the most important (and difficult) step when trying to solve an information retrieval task?

Deep Learning (Neural Networks)

How does a feed forward neural network (FFNN) compute the prediction for a new data point?

How could a multi-layer FFNN be simplified, if it did not contain any non-linear activation functions between its layers?

In what way could you manipulate the parameters (i.e., weight matrices) of an existing FFNN without changing its predictions?

What type of neural network architecture would be a natural choice for sequential data like text or time series data?

What type of neural network architecture would be a natural choice for image data?

How does self-supervised learning work (e.g., using text data)?

How does transfer learning work and when can it help?

Time Series Forecasting

What kind of input features can you use in a time series forecasting problem?

What conditions need to be fulfilled so it makes sense to use a stateless time series forecasting model?

Recommender Systems

What kind of problems (in terms of inputs and outputs) can you solve with recommender systems?

What is the “cold start problem” and how could you circumvent it?

Avoiding Common Pitfalls

What are some things you can do to make sure the learned model is not completely wrong?

What is the difference between data and concept drift?

Interpolation: Does the model generalize?

How can you determine if a model over- or underfits the data?

What can you do to improve the performance in case of <u>underfitting</u>?

What can you do to improve the performance in case of <u>overfitting</u>?

Why does the performance on the training set get worse as the size of the training set increases?

Why should you not use a univariate feature selection approach? What are better alternatives?

Why can the performance get worse if you (aggressively) reduce the dimensionality of the data with PCA?

Extrapolation: Correlation vs. Causation

Why do ML models often fail to extrapolate, i.e., do not make reliable predictions for data points outside the training domain?

What are “Adversarial Attacks”?

What can you do to try to catch and prevent systematic bias?

Explainability & Interpretable ML

What is the difference between local and global explainability?

Name two intrinsically interpretable models:

-
-

How can you explain an individual prediction of a linear model?

How is the permutation feature importance computed?

How is a partial dependence plot generated?

How can an intrinsically interpretable surrogate model be used to explain an individual prediction of a more complex model?

How can you generate optimal inputs and counterfactual examples for a neural network (e.g., for adversarial attacks)?

Reinforcement Learning

For which kinds of tasks does it make sense to use reinforcement learning and when does a normal optimization suffice?

How does the Epsilon-Greedy Policy manage the trade-off between exploration and exploitation?

How does Q-learning for tabular RL work?

How can Q-learning be extended to work with an infinite number of states?

Which factors can complicate the use of reinforcement learning?

Conclusion

AI Transformation of a Company

What can you do if you have "big data"?

According to Andrew Ng, what are the 5 steps for a successful AI transformation of a company?

1.
2.
3.
4.
5.