Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Data:

- Where will you get data points?
- What will the features be?
- How will you need to process data?

Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Data:

- Where will you get data points?
- What will the features be?
- How will you need to process data?

Model:

- What algorithm(s) will you use?
- What hyperparameters will be important?
- How will you train it?

Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Data:

- Where will you get data points?
- What will the features be?
- How will you need to process data?

Model:

- What algorithm(s) will you use?
- What hyperparameters will be important?
- How will you train it?

Evaluation:

- What will be the loss function?
- How will hyperparameters be tuned?
- How will you assess IID and OOD performance?

Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Data:

- Where will you get data points?
- What will the features be?
- How will you need to process data?

Model:

- What algorithm(s) will you use?
- What hyperparameters will be important?
- How will you train it?

Evaluation:

- What will be the loss function?
- How will hyperparameters be tuned?
- How will you assess IID and OOD performance?

Interpretation:

- Intrinsic or post hoc?
- Local or global explanations?
- Understanding model or generalization?

Navigating an ML project

CEMS
Chemical Engineering
& Materials Science

Problem statement:

- What do you want to predict or learn?
- What would this enable?
- How do you define success?

Data:

- Where will you get data points?
- What will the features be?
- How will you need to process data?

Model:

- What algorithm(s) will you use?
- What hyperparameters will be important?
- How will you train it?

Evaluation:

- What will be the loss function?
- How will hyperparameters be tuned?
- How will you assess IID and OOD performance?

Interpretation:

- Intrinsic or post hoc?
- Local or global explanations?
- Understanding model or generalization?

Prediction:

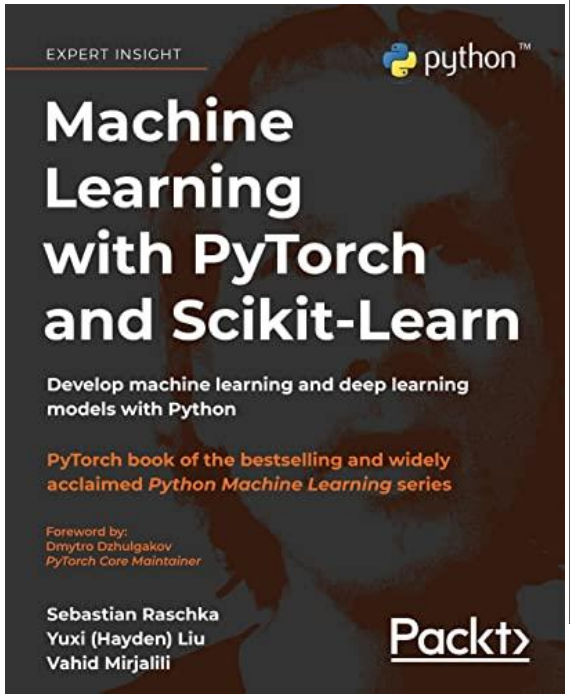
- How should this model be used?
- What are the limitations?
- Where is the model/data/code?

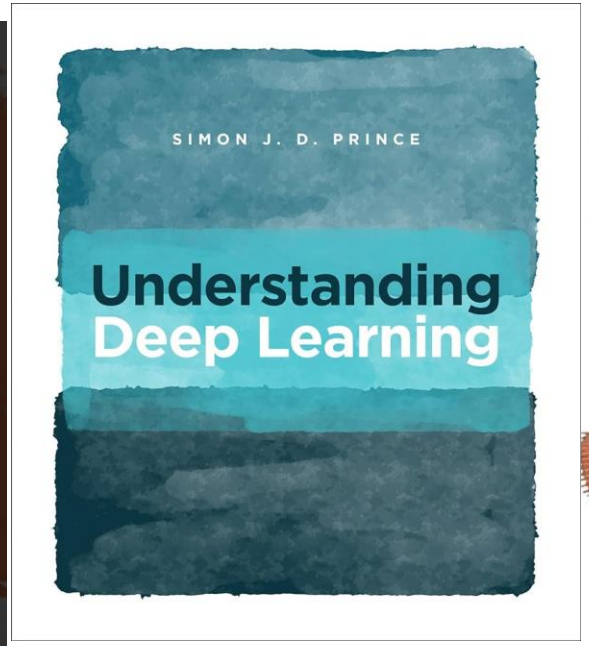
Shallow vs deep learning

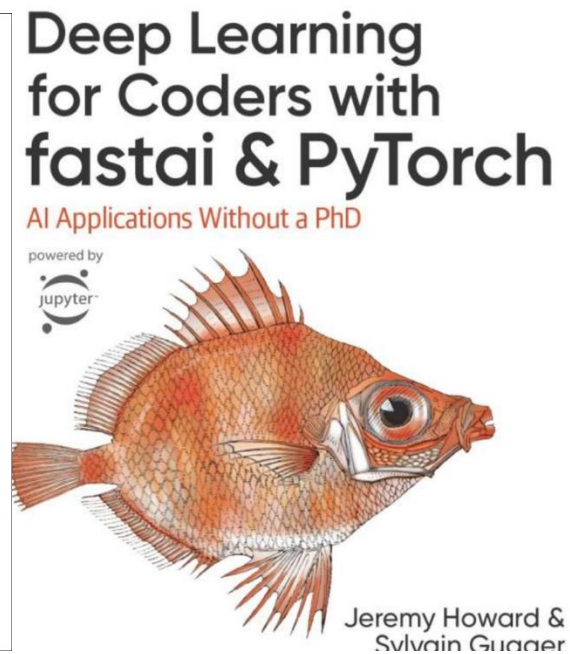
CEMS
Chemical Engineering
& Materials Science

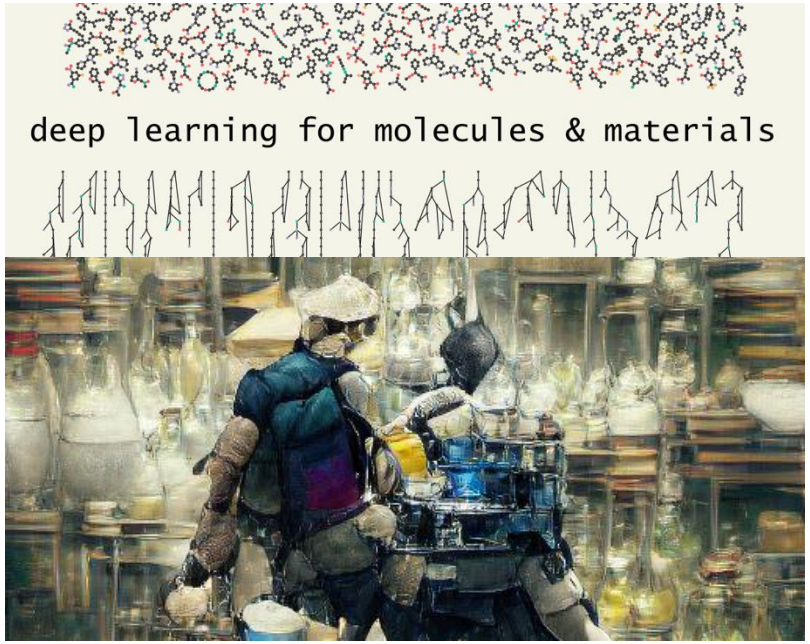
Where to go next

CEMS
Chemical Engineering
& Materials Science





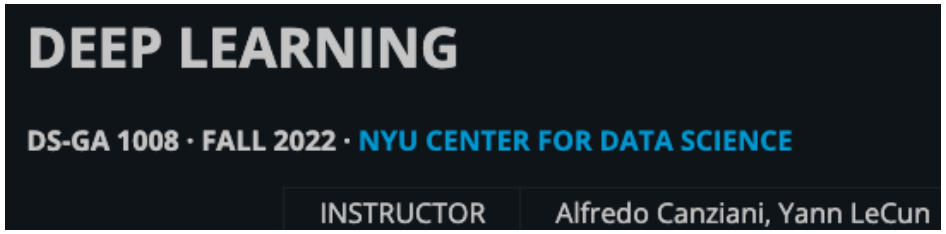




CS231n: Deep Learning for Computer Vision

YouTube

Stanford - Spring 2023



Some quick advice on coding

CEMS
Chemical Engineering
& Materials Science

- Use .py instead of .ipynb files

Some quick advice on coding

CEMS
Chemical Engineering
& Materials Science

- Use .py instead of .ipynb files

- **Everything should be in a function**

Some quick advice on coding

CEMS
Chemical Engineering
& Materials Science

- Use .py instead of .ipynb files
- Everything should be in a function
- **Build classes to collect related functions/properties**

Some quick advice on coding

CEMS
Chemical Engineering
& Materials Science

- Use .py instead of .ipynb files
- Everything should be in a function
- Build classes to collect related functions/properties
- **Use an IDE (e.g., VSCode+CoPilot)**

Some quick advice on coding

CEMS
Chemical Engineering
& Materials Science

- Use .py instead of .ipynb files
- Everything should be in a function
- Build classes to collect related functions/properties
- Use an IDE + AI assistant (e.g., VSCode+CoPilot)
- **Make sure your work is reproducible (track experiments, build python packages, use environments)**

Using ML for your research

CEMS
Chemical Engineering
& Materials Science

1) **Let your fundamental research questions drive the use of ML**

Using ML for your research

CEMS
Chemical Engineering
& Materials Science

1) Let your fundamental research questions drive the use of ML

2) **You are (probably) not going to invent a new ML method from scratch**

Using ML for your research

CEMS
Chemical Engineering
& Materials Science

1) Let your fundamental research questions drive the use of ML

2) You are (probably) not going to invent a new ML method from scratch

3) **Still, your domain knowledge can lead to creative data-driven solutions and solve your problems!**

Most important things to think about

CEMS
Chemical Engineering
& Materials Science

1) Do I have enough data to train an ML model?

Most important things to think about

CEMS
Chemical Engineering
& Materials Science

1) Do I have enough data to train an ML model?

2) **Do I understand the problem well enough to propose good features?**

Most important things to think about

CEMS
Chemical Engineering
& Materials Science

1) Do I have enough data to train an ML model?

2) Do I understand the problem well enough to propose good features?

3) **Does my validation/testing scheme align with how the model will be used in practice?**

Most important things to think about

CEMS
Chemical Engineering
& Materials Science

1) Do I have enough data to train an ML model?

2) Do I understand the problem well enough to propose good features?

3) Does my validation/testing scheme align with how the model will be used in practice?

4) Do my results pass a sanity check?!