

Assigned videos and readings

1. Computational thinking

- a. Introduction [5 min]: <https://www.youtube.com/watch?v=qbnTZCj0ugI>
- b. Introduction [Reading]: <https://www.americanscientist.org/article/computational-thinking-in-science>
- c. A perspective on programming vs coding [first 3 min]: <https://www.youtube.com/watch?v=rkZzg7Vowao>
- d. Be curious [11 min]: <https://www.youtube.com/watch?v=iUbd7EoQnK>
- e. Solving a first problem
 - i. [40 min]: <https://www.youtube.com/watch?v=C1lhuz6pZC0>
 - ii. [40 min]: <https://www.youtube.com/watch?v=uK5yvoXnkSk>

2. Statistics & probability

- a. Stochastic thinking [40 min]: <https://www.youtube.com/watch?v=-1BnXEwHUok>
- b. Sampling [40 min]: https://www.youtube.com/watch?v=soZv_KKax3E
- c. Hypothesis testing [15 min]: <https://www.youtube.com/watch?v=0oc49DyA3hU>
- d. P-value caution [11 min]: <https://www.youtube.com/watch?v=vemZtEM63GY>
- e. Maximum likelihood [6 min]: <https://www.youtube.com/watch?v=XepXtl9YKwc>
- f. Extra reading:
 - i. Probability concepts and terminology: https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/mit18_05_s22_probability_pdf/
 - ii. Statistics concepts and terminology: https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/mit18_05_s22_statistics_pdf/

3. Data handling & visualization

- a. Exploratory data analysis with Python [40 min]: <https://www.youtube.com/watch?v=xi0vhXFegw>
- b. Basic visualization with Python [35 min]: <https://www.youtube.com/watch?v=UO98IJQ3QGI>
- c. Story telling with graphs design [50 min]: <https://www.youtube.com/watch?v=Hfx1X9WSGYQ>

4. Dimensionality reduction

- a. Introduction [first 24 min]: <https://www.youtube.com/watch?v=pAwjiGkafBM>
- b. More technical [first 33 min]: <https://www.youtube.com/watch?v=MnRskV3NY1k>
- c. Principal component analysis
 - i. Main idea [22 min]: <https://www.youtube.com/watch?v=FgakZw6K1QQ>
 - ii. Practical points [8 min]: <https://www.youtube.com/watch?v=oRvgq966yZg>
 - iii. Parallel explanation with Python code [7 min]: <https://www.youtube.com/watch?v=Oi4SJqJIL2E>
 - iv. Extra: https://www.youtube.com/watch?v=xBf_LZ5ZgY4
- d. t -distributed stochastic neighbor embedding [12 min]: <https://www.youtube.com/watch?v=NEaUSP4YerM>

5. Clustering

- a. Introduction [50 min]: <https://www.youtube.com/watch?v=esmzYhuFnds>
- b. k-mean clustering [8 min]: <https://www.youtube.com/watch?v=4b5d3muPQmA>

- c. hierarchical clustering [11 min]: <https://www.youtube.com/watch?v=7xHsRkOdVwo>
- d. DBSCAN [9 min]: <https://www.youtube.com/watch?v=RDZUdRSDOOK>
- e. More technical [53 min]: <https://www.youtube.com/watch?v=mU3GZaOoVDA>

6. Machine learning

- a. Introduction & some models
 - i. [50 min]: <https://www.youtube.com/watch?v=h0e2HAPTGF4>
 - ii. [50 min]: <https://www.youtube.com/watch?v=eg8DJYwdMyg>
- b. Linear model
 - i. Ridge regression [20 min]: <https://www.youtube.com/watch?v=Q81RR3yKn30>
 - ii. LASSO regression [8 min]: <https://www.youtube.com/watch?v=NGf0voTMIcs>
 - iii. Logistic regression
 - 1. [19 min]: <https://www.youtube.com/watch?v=vN5cNN2-HWE>
 - 2. [11 min]: <https://www.youtube.com/watch?v=BfKanl1aSG0>
 - iv. Support vector machine [20 min]:
<https://www.youtube.com/watch?v=efR1C6CvhmE>

7. Tree model

- a. Decision tree [18 min]: <https://www.youtube.com/watch?v=L39rN6gz7Y>
- b. Random forest [10 min]: https://www.youtube.com/watch?v=J4Wdy0Wc_xQ
- c. Adaptive boosting [21 min]: <https://www.youtube.com/watch?v=LsK-xG1cLYA>
- d. Bias versus variance [6 min]: <https://www.youtube.com/watch?v=EuBBz3bl-aA>
- e. XGBoost in Python [56 min]: <https://www.youtube.com/watch?v=GrJP9FLV3FE>

8. AI & deep learning

- a. Introduction to AI [60 min, skip a lot]: <https://www.youtube.com/watch?v=16Dir4QqCUg>
- b. Deep learning [58 min]: <https://www.youtube.com/watch?v=QDX-1M5Nj7s>
- c. Convolutional neural networks [55 min]:
https://www.youtube.com/watch?v=NmLK_WQBxB4

9. Explainability

- a. Motivation [28 min]: https://www.youtube.com/watch?v=DYQdP_F-LA
- b. Techniques [first 30 min]: <https://www.youtube.com/watch?v=6n8r523QP8>
- c. Evaluation [first 40 min]: <https://www.youtube.com/watch?v=htjpbvHJQo>

10. AI project design

- a. Medical AI consideration [48 min] <https://www.youtube.com/watch?v=UZEstizNxkg>
- b. ML product decision [31 min]: <https://www.youtube.com/watch?v=2aWh3-Wnb-A>

More materials

- MIT 6.S191 deep learning & AI: https://www.youtube.com/playlist?list=PLtBw6njQRU-rwp5_7C0oIVt26ZgjG9NI
- Tubingen intro to ML:
<https://www.youtube.com/playlist?list=PL05umP7R6ij35ShKLDqccJSDntugY4FQT>
- MIT 6.0002 computational thinking:
https://www.youtube.com/playlist?list=PLUI4u3cNGP619EG1wp0kT-7rDE_Az5TND
- MIT 6.S897 ML for healthcare: <https://ocw.mit.edu/courses/6-s897-machine-learning-for-healthcare-spring-2019/>
- StatQuest YouTube channel: <https://www.youtube.com/@statquest>

Assigned Python practices

Feel free to find interesting problem to practice on <https://leetcode.com/problemset/> and <https://programming.in.th/tasks>

Week 1

- Python editors [25 min]: https://www.youtube.com/watch?v=5pf0_bpNbkw
- <https://www.kaggle.com/learn/intro-to-programming>
- <https://www.kaggle.com/learn/python>

Week 2

- <https://www.kaggle.com/learn/pandas>
- <https://www.kaggle.com/learn/data-visualization>
- Kaggle Titanic dataset: <https://www.kaggle.com/competitions/titanic>
 - Explore the distribution of passenger characteristics
 - Identify characteristics that contribute to survival with statistical tests
 - Tell a story with graphs

Week 3

- <https://www.kaggle.com/learn/data-cleaning>
- https://www.pythonhealthdatascience.com/content/02_stat_prog/03_exercises_front_page.html
 - ED data wrangling
 - Stroke data wrangling
 - Visualizing time-series data
- Operating room utilization dataset: https://www.dropbox.com/scl/fi/rrua76vmq6m224awlunby/2022_Q1_OR_Utilization.xlsx?rlkey=6vecn2leeh1a0ufh5wzpq4sgn&dl=0
 - Perform data exploration and visualization
 - Tell a story from the data
 - Study queue simulation [22 min]: <https://www.youtube.com/watch?v=QppldN-t4pQ>
 - Implement a queue simulation from OR data
 - Simulate some synthetic datasets
- Colorectal dataset
 - Perform various dimensionality reduction and clustering techniques to identify subtypes
 - Compare the results
 - Identify genes and mutations associated with each subtypes
 - Tell a story about your finding

Week 4

- <https://www.kaggle.com/learn/intro-to-machine-learning>
- <https://www.kaggle.com/learn/intermediate-machine-learning>

- <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016>
 - Predicting hospital admission at emergency department

Week 5

- <https://www.kaggle.com/learn/intro-to-deep-learning>
- <https://www.kaggle.com/learn/computer-vision>
- Kaggle digit MNIST: <https://www.kaggle.com/competitions/digit-recognizer>
 - Perform dimensionality reduction
 - Use classical techniques
 - Implement a fully-connected network
 - Implement a CNN
 - Use a pre-trained CNN
- <https://github.com/cmb-chula/cu-covid19-isolation/tree/main>
 - Predict hospital admission following COVID-19 home isolation

Week 6

- <https://www.kaggle.com/learn/time-series>
- <https://www.kaggle.com/learn/machine-learning-explainability>
- https://colab.research.google.com/github/alistairewj/cvpr-2021-cxr-tutorial/blob/main/cvpr_2021_tutorial.ipynb
 - Just study and run the code and we will talk about it