[SHA48] INFORMATION THEORY

- Highlights
  - Consider the problem of describing the nature of information in a quantitative way -- to measure the amount of information. Or when you have to wrap up everything in a digital format, how *many* “basic information units” are needed to describe a system of interest? How is the number related to the nature of the system, and what is this nature?
- Further Reading
  - A decision tree algorithm family — link to a tutorial.
  - A online course on information theory.

[COR95] SUPPORT VECTOR MACHINES

- Highlights
  - Support vector machine (SVM) is motivated by the authors' early work investigating the answer to the following fundamental question:

    Why can we build a system based on empirical experience (simply, training data consisting of system input and desired output in the context of some task) and expect the system to work (almost) as reliably as one built on the true laws governing the environment where the task takes place?

    Of course, the true laws are unknown in practice. Otherwise, it will defeat the purpose of machine learning.
- Further Reading
  - The "Learning from Data" book comes with a series of video lectures. Starting from Lecture 14, Prof. Abu-Mostafa introduced SVM. You may need to "back-track" multiple video lectures to understand what is going on L14, though.
  - Christopher J.C. Burges published a high-quality comprehensive introduction to SVM in 1998, which is friendly to engineers.

[COO90] INFERENCE IN BAYESIAN NETS IS HARD

- Highlights

  Bayesian networks make an elegant family of data models that represents dependencies in variables using *graphs*. Therefore it allows one to analyse (normally considered "algebraic") properties of a data model using graph theory. But when trying to reach some conclusion from observed data to the unknown target, the computational tasks may be overwhelming. This paper is highly technical, check the further reading section for some comprehensive introduction.

- Further Reading

  - Zoubin Ghahramani did a good tutorial to Bayesian inference and graphical models.

  - Judea Pearl (one of the founders of Bayes Nets) published a comprehensive overview.

[FRI01] GREEDY BOOSTING

- Highlights
  - It is possible to use the sum of simple functions to approximate complicated functional behaviour given training data.   The incremental steps are made easier (than a holistic function adjust) by the proposed technique.
- Further Reading
  - Please explore by searching using keywords such as “XGBoost”
  - You may want to explore winning entries to some Kaggle competitions.

[IND99] APPROXIMATE NEAREST NEIGHBOUR

- Highlights
    - It represents the long-time effort to render the data samples in a way that facilitates "reference checking" for test data. The research area represented by this work puts the focus on the efficiency of the expected "reference checking".
- Further Reading
    - Alexandr Andoni and Piotr Indyk wrote a review paper in 2008. They mentioned their own work as well as compared relevant works in context.
    - Muja and Flann in 2009 developed an algorithm that is widely adopted in computer vision projects. You may need to know something about the algorithm if interested in building robots that can move around by seeing the world as a human does.

[BEL97] PCA / LDA IN IMAGE PROCESSING

- Highlights
  - As above, PCA/LDA represents another aspect of improving “to refer to the seen (training) data when facing unseens (test)”. Such methods look for a transformation of the data into some space, in which the distance is more ”meaningful”, i.e. facilitates solving the prediction task.
- Further Reading
  - Note LDA dates back to a classic work by Fisher in 1936, you may also review that paper. (Caveat: symbols/denotations could look obscure.)
  - Martinez and Kak did a critical study of the area in 2000.
  - In a broader sense, data representation learning is literally the core of machine learning. The searching for good representations leads to modern deep learning.

[RUM86] BACKPROP GRADIENT COMPUTATION

- Highlights
  - It addresses the difficulty of figuring out how to adjust model parameters, which are organised in complicated structures. And the parameters contribute to the model output via involved computations. It is the bedrock of modern neural network training algorithms.
- Further Reading
  - You can find back-propagation in almost all deep learning techniques. To be more specific, you can check out the adaptive gradient optimisation algorithm developed by Kingma and Jimmy Lei Ba in 2015.
  - Recently, Hinton's group proposed another optimisation technique, interestingly named "Look Ahead".

[LEC98] CONVOLUTIONAL NEURAL NETWORK

- Highlights
  - Convolutional neural networks are arguably the most applied neural network architecture in state-of-the-art deep learning technologies. It makes it feasible to build a decent level of sophistication in computational models when the data have a large number of attributes.
- Further Reading
  - Convolutional network family is the workhorse of modern computer vision. Several important developments are widely used behind many amazing artificial intelligence applications, including the first modern practical "deep" net developed by Alex Krizhevsky (with Hinton), VGG ("very deep") by Simonyan and Zisserman, Inception nets (multiple versions, evolving) by Google and ResNet (extra step to reduce model complexity) by Kaiming He et al. Read this blog for an assessable overview.
  - More broadly, there are some recent efforts to take advantage of internal interactions between the attributes of data (So computational models can be both powerful and less complex.) The focus is on data with attributes organised as graphs such as social media data or data representing interactions between real-world objects or people. Check out "graph convolutional neural networks" if interested.

[HOC97] LONG-SHORT TERM MEMORY FOR RECURRENT NEURAL NETWORKS

- Highlights
  - Recurrent neural networks (RNN) represent a family of data models with "states". When processing data, the model keeps a working memory of what it has seen. This work helps train neural networks with learnable working memory units.
- Further Reading
  - RNN family as a powerful computing model with memory can be very useful in a lot of application areas. Here is a nice comprehensive RNN roadmap with applications.
  - One traditional strength of RNN is its capability of processing natural languages since the tokens (minimum semantic units, consider words for English) of a language makes sense when they are organised in particular sequences. To process the sequences you need to keep a working memory about what you have seen. In recent years, a new type of self-reference model has been proposed to address the need for modelling internal relationships in data attributes. Please explore the development of "attention models", if you are interested in natural language processing for tasks such as machine translation or chat robots. Here is a good introduction to one of the foundation works. **Were there a bit critical comments, it would also be an A+ report of literature reading.**

[GOO14] GENERATIVE MODELLING

- Highlights
  - GAN is an effective representation method of data distribution. This model is distinctive in the way it is being used: it produces samples from the estimated data distribution from *random noises*.
- Further Reading
  - Hong, Hwang, Yoo, and Yoon have been maintaining a comprehensive overview of both theoretical and practical development of the GAN models since 2017. Their overview paper has been updated in Feb 2019.
  - Get to the nature of how a GAN model works, they can be used in amazingly creative ways. You can check out some recent developments by NVIDIA or Deepmind for using GAN models to generate impressive "counterfeit" examples, including people portrait, artworks etc.