

# H6751 Summary

Zhao Rui

# Agenda

1. Modern AI
2. Course Summary

# Modern AI

# Modern AI (90s-present)

- **Stat Model:** Pearl (1988) promote Bayesian networks in AI to **model uncertainty** (based on Bayes rule from 1700)

**Stat Model:** infer the relationship among variable in data

- **Machine Learning:** Vapnik (1955) invented support vector machines to **learn parameters** (based on statistical models in early 1900s)

**Machine Learning:** sacrifice interpretability for predictive power

<https://www.nature.com/articles/nmeth.4642>

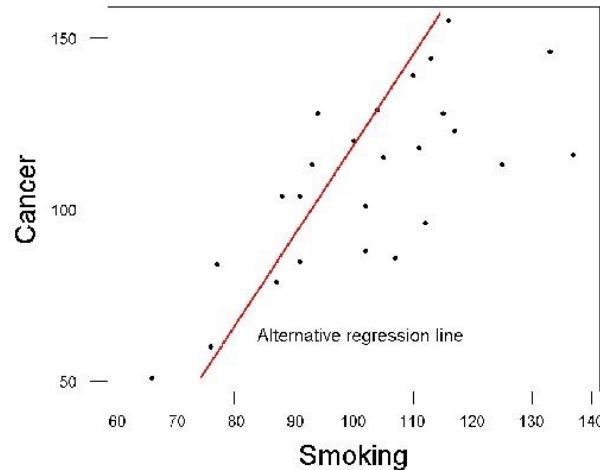
# Take Linear Regression as the example

## Stat Model:

1. **Inference:** Characterize the relationship between the smoking index and cancer rates.
2. Conduct the significance test of the model parameters

## ML:

1. **Prediction:** Get a model that is able to make prediction of the cancer rates based on smoking index
2. Evaluate the model performance over testing data.



# Course Summary

Quant investing

+ Add to myFT

# Glitchy coronavirus markets cause quant funds to misfire

Renaissance, Two Sigma and DE Shaw suffer unusual setbacks



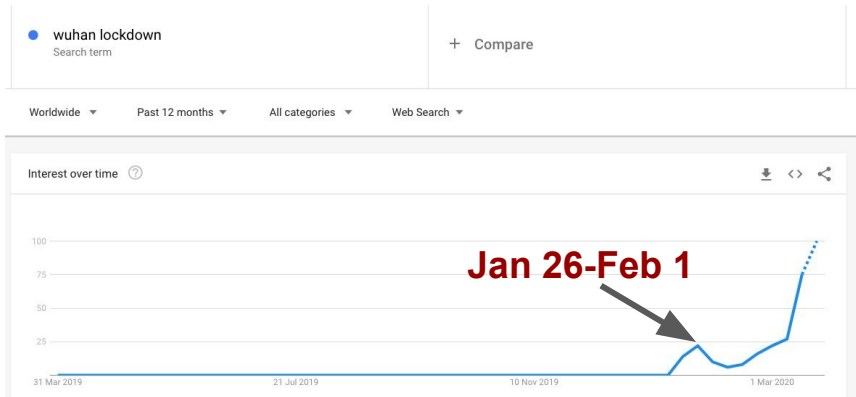
# Overfitting

- The common practice in quant research: after conducting **hundreds** or even **thousands times** backtesting, the best strategy (highest sharpe ratio) is selected.
  - Selection bias
  - Testing data or out-of-sampled data is **misused** as validation data
  - Overfitting!!!
- In hypothesis test, the testing is used to **refute** a false claim instead of building a claim
- **Explainability** matters (Try to build theories, not a complex and black box)



# Prediction

- Sell-off is the black swan to Quant models based on history prices or fundamental data or cross-sectional factors
  - The future trend is unpredictable
- However, it is possible to find hidden states behind huge amounts of unstructured data
  - How to filter noise (statistical hypothesis testing)



Investing

- Three Main Topics:
  - Text Pre-processing Techniques
  - Text classification (Data Mining Models)
  - Deep Learning for Text data
- How do we understand the concepts of machine learning models better:
  - Build your own knowledge graph that can explain the connections among all these models
  - Check its corresponding applications

| Date             | Topic                             |
|------------------|-----------------------------------|
| Sat a.m<br>01/18 | Introduction to Text Mining       |
| Sat a.m<br>02/01 | Pre-processing for Text Mining I  |
| Sat p.m<br>02/01 | Pre-processing for Text Mining II |
| Sat a.m<br>02/15 | Text Categorization I             |
| Sat p.m<br>02/15 | Text Categorization II            |
| Sat a.m<br>02/29 | Text Categorization III           |
| Sat p.m<br>02/29 | Document Clustering               |
| Sat a.m<br>03/21 | Sentiment Analysis                |
| Sat p.m<br>03/21 | Introduction to Deep Learning     |
| Sat a.m<br>04/04 | Word Embeddings                   |
| Sat p.m<br>04/04 | Recurrent Neural Network          |
| Sat a.m<br>04/18 | Convolutional Neural Network      |

***There is the possibility that people will organize, become engaged, as many are doing, and bring about a much better world, which will also confront the enormous problems, that we're facing right down the road***

by Noam Chomsky

