

What is data
science?

What is data science?

- Answer scientific questions by (or at least with the help of) data

Example: 台中市空氣品質

- 台中市空氣品質是否改善？
 - Data available [here](#)
- 統計/資料分析

Example: 自駕車

- [How Data Science Is Driving The Driverless Car](#)
 - [BDD100K](#) dataset by UC Berkeley
- 機器學習/AI, big data

Example: 天氣預報

- Numerical weather model + observational data
- [The data science behind weather prediction](#)
- 資料同化(data assimilation), big data

子領域

1. Statistics
2. Machine learning
3. Big data
4. Data assimilation

Statistics

- 以管窺天: explore the properties of a population by a (usually small and random) sample
- Focus on scientific insights (inferencing)
- Prefer interpretable models
- Results are usually reliable if you do it correctly

Machine learning

- Prediction, prediction, prediction!
- Find a model to predict future events
- Generalization may be an issue due to some careless mistakes (e.g., data collection)

ML v.s. Stat

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = \$1,000,000	large grant = \$50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

Data assimilation

- Putting scientific models and observations together
 - Estimate unknown parameters in a scientific model
 - Correct forecasts by observations
 - Extrapolate observations with the help of scientific (mathematical) models
- Closely related to all the above three fields

Data science pipeline

