Unsupervised Analysis: Best Practices

Exploratory vs. Confirmatory Analysis

Confirmatory Analysis:

- Seeks to test an a priori hypothesis.
- Examples:
  - Classical inferential statistics.
  - Prediction in supervised learning.

Exploratory Analysis:

- Seeks to make data-driven discoveries.
- Hypothesis generating.

Which is unsupervised analysis?

Validating Data-Driven Discoveries

- Corroborate via existing literature.
- Show data-driven discovery is stable.
  - Small changes to the data, the algorithm, the method, the parameters, etc. yield the same result.
  - Multiple approaches yield the same result.
- Validate via biological experiments.
  - True confirmation.
  - Expensive & sometimes not possible.
- Confirm via a completely separate test set.

Confirming Discoveries on a Test Set

Exercise

Suppose you use a training data set to BLANK. How would you use a separate test data set to validate this discovery?

BLANK:

- Discover a major pattern.
- Discover clusters.
- Discover important features.
- Discover important connections between features.

Some Good Rules for Unsupervised Analysis

1. Always visualize.
2. Use multiple techniques.
3. Validate discoveries when possible.
4. Communicate uncertainty.
5. Make your analysis reproducible.