Just like in econometrics, the most basic "learning algorithm" is OLS

Just like in econometrics, the most basic "learning algorithm" is OLS

Or, if Y is categorical, logistic regression

Just like in econometrics, the most basic "learning algorithm" is OLS

Or, if Y is categorical, logistic regression

But, we know there are many other ways to estimate models

Just like in econometrics, the most basic "learning algorithm" is OLS

Or, if Y is categorical, logistic regression

But, we know there are many other ways to estimate models

e.g. non-parametric, semi-parametric, Bayesian, ...

- Supervised learning: we predict Y from a set of X's

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Most of econometrics is about supervised learning (i.e. estimate $\hat{\beta}$)

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Most of econometrics is about supervised learning (i.e. estimate $\hat{\beta}$)

- But there are some elements of unsupervised learning

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Most of econometrics is about supervised learning (i.e. estimate $\hat{\beta}$)

- But there are some elements of unsupervised learning

  - Particularly with regards to detecting unobserved heterogeneity types

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Most of econometrics is about supervised learning (i.e. estimate $\hat{\beta}$)

- But there are some elements of unsupervised learning

  - Particularly with regards to detecting unobserved heterogeneity types

  - e.g. factor analysis (detect types based on a set of measurements)

- Supervised learning: we predict Y from a set of X's

- Unsupervised learning: we try to group observations by their X's

- Most of econometrics is about supervised learning (i.e. estimate $\hat{\beta}$)

- But there are some elements of unsupervised learning

    - Particularly with regards to detecting unobserved heterogeneity types

    - e.g. factor analysis (detect types based on a set of measurements)

    - e.g. the EM algorithm (detect types based on serial correlation of residuals)

Examples of supervised learning algorithms:

Examples of supervised learning algorithms:

- Tree models

Examples of supervised learning algorithms:

- Tree models
  - Basically a fully non-parametric bin estimator

Examples of supervised learning algorithms:

- Tree models
    - Basically a fully non-parametric bin estimator
    - Can generalize to "forests" that average over many "trees"

Examples of supervised learning algorithms:

- Tree models
    - Basically a fully non-parametric bin estimator
    - Can generalize to "forests" that average over many "trees"

- Neural networks

Examples of supervised learning algorithms:

- Tree models

  - Basically a fully non-parametric bin estimator

  - Can generalize to "forests" that average over many "trees"

- Neural networks

  - Model the human brain's system of axons and dendrites

Examples of supervised learning algorithms:

- Tree models

  - Basically a fully non-parametric bin estimator

  - Can generalize to "forests" that average over many "trees"

- Neural networks

  - Model the human brain's system of axons and dendrites

  - "Input layer" is the X's, "Output layer" is Y

Examples of supervised learning algorithms:

- Tree models

    - Basically a fully non-parametric bin estimator

    - Can generalize to "forests" that average over many "trees"

- Neural networks

    - Model the human brain's system of axons and dendrites

    - "Input layer" is the X's, "Output layer" is Y

    - "Hidden layers" nonlinearly map the input and output layers

Examples of supervised learning algorithms:

- Tree models

    - Basically a fully non-parametric bin estimator

    - Can generalize to "forests" that average over many "trees"

- Neural networks

    - Model the human brain's system of axons and dendrites

    - "Input layer" is the X's, "Output layer" is Y

    - "Hidden layers" nonlinearly map the input and output layers

    - Mapping ends up looking like a logit of logit of logits

Examples of supervised learning algorithms:

- Tree models

  - Basically a fully non-parametric bin estimator

  - Can generalize to "forests" that average over many "trees"

- Neural networks

  - Model the human brain's system of axons and dendrites

  - "Input layer" is the X's, "Output layer" is Y

  - "Hidden layers" nonlinearly map the input and output layers

  - Mapping ends up looking like a logit of logit of logits

  - Linear regression is a special case

- Bayesian models

- Bayesian models
  - Bayes' rule can be thought of as a learning algorithm

- Bayesian models
    - Bayes' rule can be thought of as a learning algorithm
    - Use it to update one's prior

- Bayesian models
    - Bayes' rule can be thought of as a learning algorithm
    - Use it to update one's prior

- Support Vector Machine (SVM)

- Bayesian models
  - Bayes' rule can be thought of as a learning algorithm
  - Use it to update one's prior

- Support Vector Machine (SVM)
  - Originally developed for classification

- Bayesian models
  - Bayes' rule can be thought of as a learning algorithm
  - Use it to update one's prior

- Support Vector Machine (SVM)
  - Originally developed for classification
  - Tries to separate 0s and 1s by as large of a margin as possible

- Bayesian models
  - Bayes' rule can be thought of as a learning algorithm
  - Use it to update one's prior

- Support Vector Machine (SVM)
  - Originally developed for classification
  - Tries to separate 0s and 1s by as large of a margin as possible
  - Based on representing examples as points in space

- Bayesian models
    - Bayes' rule can be thought of as a learning algorithm
    - Use it to update one's prior

- Support Vector Machine (SVM)
    - Originally developed for classification
    - Tries to separate 0s and 1s by as large of a margin as possible
    - Based on representing examples as points in space
    - Generalization of the maximal margin classifier

We covered the EM algorithm and PCA in previous videos

We covered the EM algorithm and PCA in previous videos

k-means clustering

We covered the EM algorithm and PCA in previous videos

k-means clustering

- Attempts to group observations together based on the X's

We covered the EM algorithm and PCA in previous videos

k-means clustering

- Attempts to group observations together based on the X's

- Choose cluster labels to minimize difference in X's among labeled observations

We covered the EM algorithm and PCA in previous videos

k-means clustering

- Attempts to group observations together based on the X's

- Choose cluster labels to minimize difference in X's among labeled observations

$$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i \in C_k} \sum_{\ell=1}^{L} (x_{i\ell} - \overline{x}_{\ell j})^2$$

We covered the EM algorithm and PCA in previous videos

k-means clustering

- Attempts to group observations together based on the X's

- Choose cluster labels to minimize difference in X's among labeled observations

$$\min_{C_1,\dots,C_K} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i \in C_k} \sum_{\ell=1}^{L} (x_{i\ell} - \overline{x}_{\ell j})^2$$

N_k is the number of observations in cluster k, L is number of X's

We covered the EM algorithm and PCA in previous videos

k-means clustering

- Attempts to group observations together based on the X's

- Choose cluster labels to minimize difference in X's among labeled observations

$$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i \in C_k} \sum_{\ell=1}^{L} (x_{i\ell} - \overline{x}_{\ell j})^2$$

N_k is the number of observations in cluster k, L is number of X's

Can choose other metrics besides Euclidean distance

- Active learning: algorithm chooses the next example it wants a label for

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Two common examples of active learning:

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Two common examples of active learning:

1. Reinforcement learning powers the world-beating chess engines

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Two common examples of active learning:

1. Reinforcement learning powers the world-beating chess engines

    - These algorithms use dynamic programming methods

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Two common examples of active learning:

1. Reinforcement learning powers the world-beating chess engines

    - These algorithms use dynamic programming methods

    - Use Conditional Choice Probabilities for computational gains

- Active learning: algorithm chooses the next example it wants a label for

- Balances "exploration" and "exploitation"

- Two common examples of active learning:

1. Reinforcement learning powers the world-beating chess engines

    - These algorithms use dynamic programming methods

    - Use Conditional Choice Probabilities for computational gains

2. Recommender systems power social networks, streaming services, etc.