

Statistical Learning Methods

Questions Lecture 1

Tobias Famos

June 18, 2022

Questions

1. What is statistical learning?
2. What is the difference between Data and Information?
3. Can you just use any data?
4. Why do we estimate $f()$?
5. How do we estimate $f()$?
6. What is the trade-off between prediction accuracy and model interpretability?
7. What is the difference between supervised and unsupervised learning?
8. What is the difference between regression and classification problems?
9. What is the difference between μ and $\hat{\mu}$?
10. What is the no free lunch theorem?

Answers

1.
 - We have data and we suppose there is a relationship between X and Y .
 - We try to model the relationship with a function $Y = f(X) + \epsilon$.
 - We model the relationship to either predict or explain the relationship.
2. Data is the encoding of information. Information is the meaning you can get out of data.
3. No you can't. Data should be preprocessed.
 - Remove outliers
 - Remove illegal values / obvious measurement errors (e.g. negative age)
 - Remove any rows with missing values
4.
 - Either to **predict** (We want to compute y for new x)
 - Or to **explain** (We want to have an explanation on how the price is influenced by different factors)
5.
 - We have a training set of data.
 - We build a model with the training set and assess it for its efficiency.
6.
 - More accurate models are more complex.
 - More complex models are more difficult to interpret.
 - Interpretability can be an important factor for model choice
 - Example: Treatment in medical setting, there it must be clear why a decision has been made.
7.
 - In supervised learning, we can observe X and Y of the training data. (e.g. linear regression)
 - In unsupervised learning we can only observe X of the training data. (e.g. clustering)
 - Thus in supervised learning, we know the classification / prediction of the elements in training data whereas in unsupervised learning we don't.

-

8. In regressions, Y is \mathbb{R} where in classification Y is an element of a finite set of classes.
9.
 - Greek letters without the hat (μ): A fixed number based on the whole population.
 - Greek letters with the hat ($\hat{\mu}$): A estimated number based on a sample.
 - The estimated values are used to infer the real values.
10. The no free lunch theorem states that no machine learning algorithm is universally better than any other. But for a specific class of problems, some algorithms can outperform others.