Seth Emery
CS5565-ISL
HW1
Jan 30, 2024

Questions

1. *What are the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification? (3 pts)*

Advantages of very flexible approaches to regression include being able to identify more complex patterns or relationships between predictors and response variables (low-bias and high variance like we discussed in class). The models may be able to be more accurate by closely fitting the data.

Disadvantages of very flexible approaches to regression are that they are prone to overfitting because they closely fit the data which makes generalization harder, also the interpretability of a very flexible model is lower because of the flexibility-interpretability trade-off discussed in class.

Advantages of very flexible approaches to classification are, like regression, that you can accurately model complex decision boundaries because the fit is higher so they can be more accurate.

Disadvantages of a very flexible approach to classification are that once again there is a higher risk of overfitting, and the interpretability of the model is lower. Instead of there being a straight line through the data with a clear boundary, it could be a curvy line that is hard to interpret

2. *Under what circumstances might a more flexible approach be preferred to a less flexible approach? (2 pts)*

Circumstances where a **more flexible approach would be preferable** to a less flexible approach include when you really want high accuracy from your model, when you have a lot of data available, and when you have non-linear relationship between the predictors and the target variable. Also when you are less concerned with interpretability.

3. *When might a less flexible approach be preferred? (2 pts)*

A less flexible approach might be preferred when you really want to be able to interpret the model, you have a small dataset, or the relationship between the predictors and the outcome is more simple.

4. *Describe the differences between a parametric and a non-parametric statistical learning approach. (3 pts)*

Seth Emery
CS5565-ISL
HW1
Jan 30, 2024

The differences between a parametric and non-parametric statistical learning approach are that parametric is usually simpler, and makes assumptions about the function form, and they might have high bias and low flexibility.
Non-parametric doesn't make as many assumptions and tries to make the function from the data. Non-parametric approaches are more flexible and have less bias and require more data.

(notes from class)
Interpretability vs. Flexibility

- Parametric is highly interpretable but moderate to low flexibility
- Non-parametric is moderate to low interpretability but highly flexible

5. *What are the advantages of a parametric approach to regression or classification (as opposed to a nonparametric approach)? (3 pts)*

Advantages of a parametric approach to regression are that they are usually simple and interpretable, they don't use as much data as non-parametric, and they use a function that can be understood

Advantages of a parametric approach to classification are mostly the same as with regression, they are simple and interpretable and can be trained with less data. They also have lower variance (at the cost of more bias)

6. *What are the disadvantages of the parametric approach? (2 pts)*

Disadvantages of a parametric approach to regression are that they have higher bias and are less flexible, can cause underfitting (less accurate)

Disadvantages of a parametric approach to classification like regression they have higher bias compared to non-parametric and they are less flexible than non-parametric.

Sources:
parametric vs non-parametric:
https://www.ibm.com/docs/en/db2woc?topic=procedures-statistics-parametric-nonparametric

flexible vs inflexible:
https://www.baeldung.com/cs/ml-flexible-and-inflexible-models

My lecture notes from class (Dr. Adu Baffour) for bias-variance trade-off and flexibility-interpretability trade-off

My lecture notes from Principles of Data Science last semester from Dr. Shah about the amount of data needed for training models

I also used the rubric for grading the assignment to try and score the max points for each category!