

Slidesmanship

Outline

- Motivation
- Model Validation Framework
- Cross Validation - What is it?
- Single Run Cross Validation
- Cross Validation for Temporal Data
- Conclusions & References

- Tell 'em what you're going to tell them
- Tell 'em
- Tell 'em what you told 'em

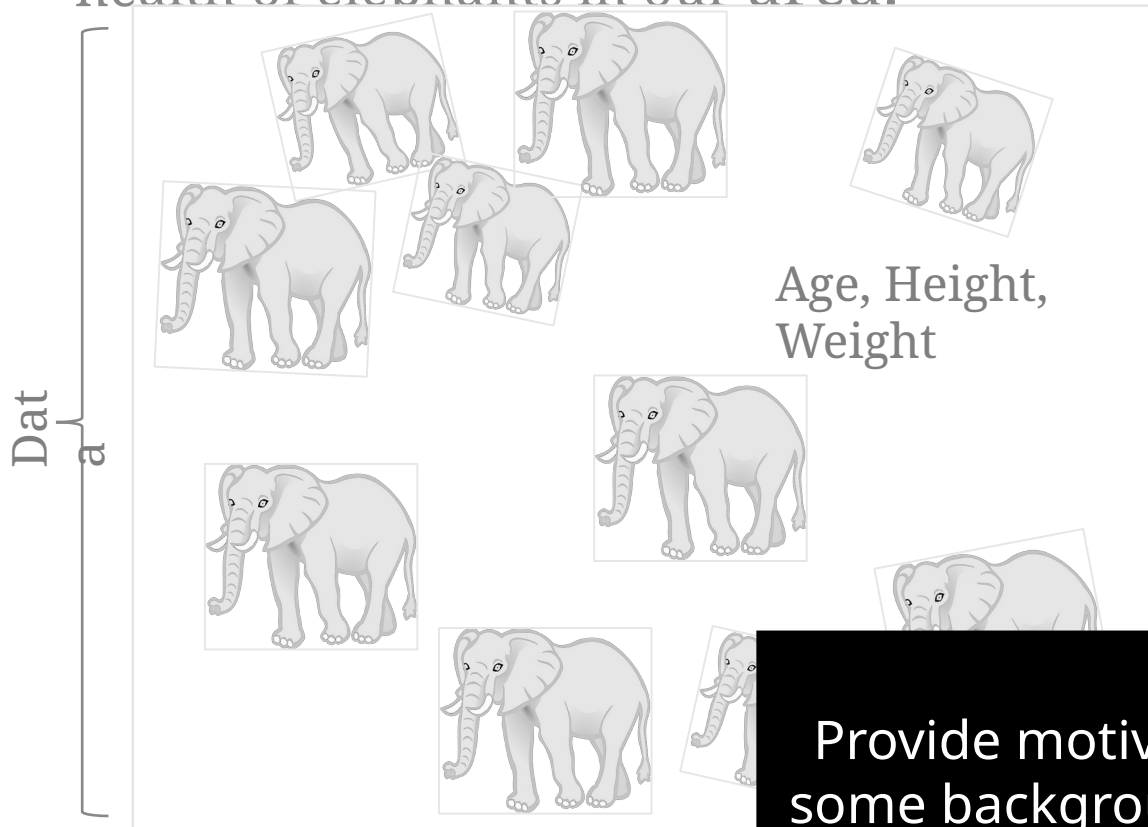
Think about what's in it for your audience. Why should they listen to you?

What should you take away?

1. What is model validation?
2. What is cross validation?
3. How do we perform cross validation for temporal data?

Motivation – Elephant Example

We want to build a model to predict an elephant's weight based on two predictors – age and height. It's difficult for us to weigh elephants so this model will be especially useful to assess the health of elephants in our area.



We will want to know how well our model will perform if we use it to predict the weight of an elephant we haven't seen

Provide motivation for your work! Give some background on the subject matter.

Our Model

Know your audience and the level of technical detail that is appropriate.

Our observation

$$y = \textit{weight} = f(\textit{height}, \textit{age}) + \epsilon$$

Our model

$$\hat{y} = \hat{f}(\textit{height}, \textit{age})$$

Goal

For new \textit{height}_n and \textit{age}_n , determine how close our prediction \hat{y}_n is to our observation y_n .

Measure of Error

Some measure of the difference between \hat{y}_n and y_n (MSE, RMSE, MAPE, etc).

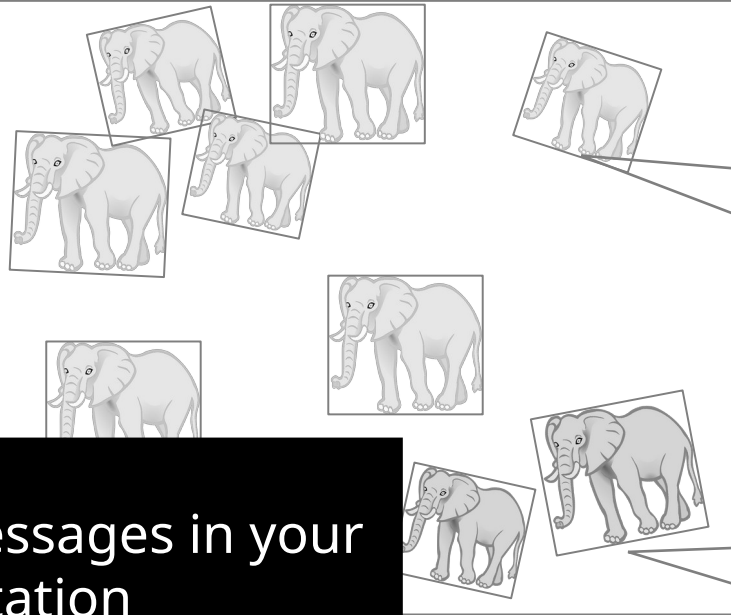
Cross Validation

To assess how our model will perform with new data, we can split our data into **training** and **testing** sets.

This is a ~~single-run~~ **cross validation**.

Training
a

Testing



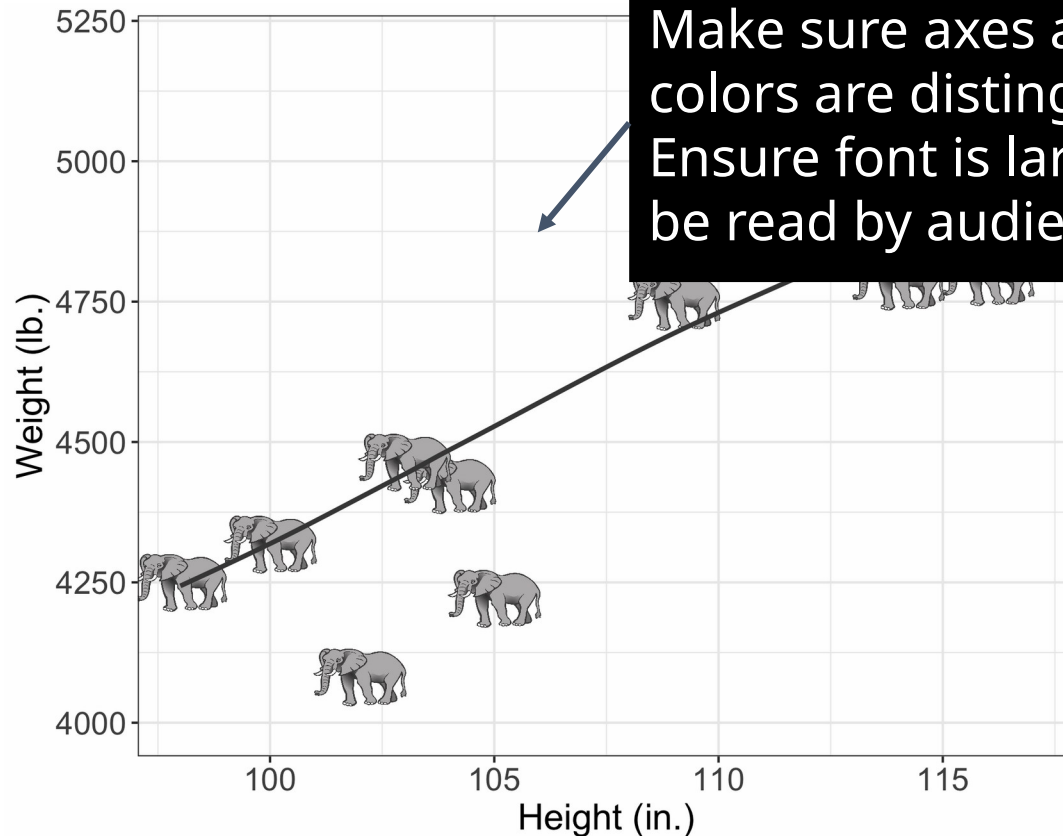
We fit our model using a **training set**.

We validate our model using a **testing set**.

Highlight key messages in your presentation

Cross validation (CV) allows us to assess our model's predictive ability using a "new" data.

Single-Run CV Limitation

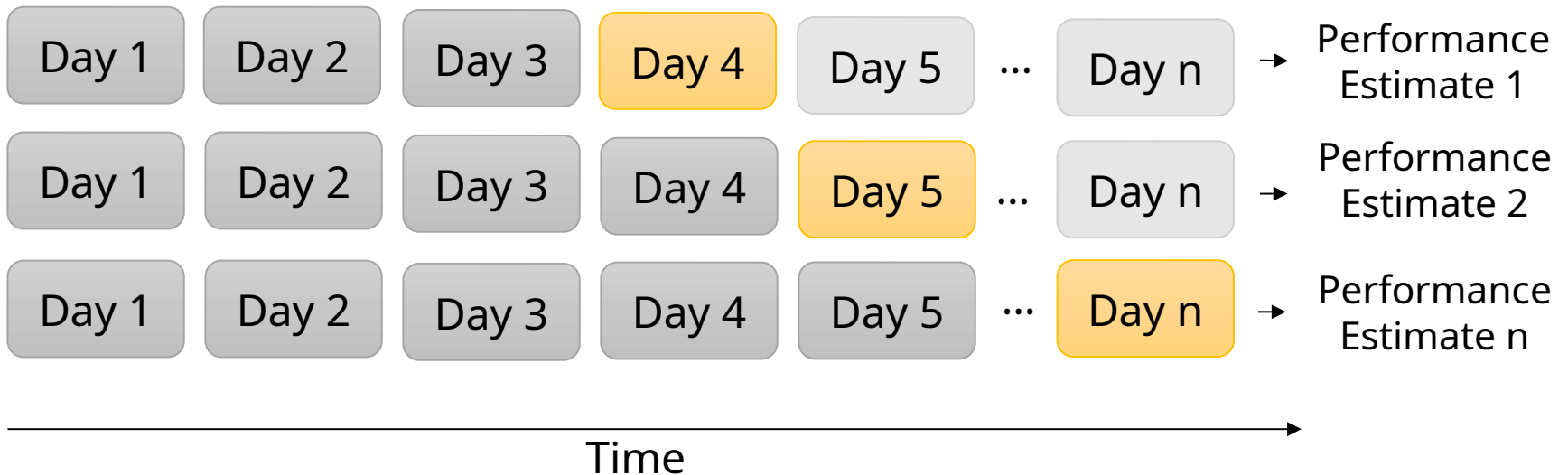


Make sure axes are labeled and colors are distinguishable. Ensure font is large enough to be read by audience.

Our model's performance may be **sensitive to the sample** that we use for our training set.

Cross Validation for **Temporal Data**

- Use color to support your message.
- Break things into steps when possible.
- Use visual aids to make complex processes more clear.



With **time series data**, we often can't split up the dataset into training and testing sets randomly.

Conclusions

- **Overfitting** occurs when your model doesn't work well on **unseen data**.
- It is important to **assess** whether your **model is overfitting** and to get **an accurate estimate of your model's performance**.
- A quick way to assess overfitting is to split your data up into training and testing sets. If your **model performs much better** for the your **training data** than your testing data, you are likely **overfitting**.
- **Cross validation** is a tool to better **assess your model's performance**.
- For **spatial** and **temporal** data, extra considerations must be taken when performing cross validation.

Tell 'em what you told 'em!

References

- Hastie, T., Tibshirani, R., Friedman, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.*
<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>
- Cross-Validation Essentials in R.
<http://www.sthda.com/english/articles/38-regression-model-validation/157-cross-validation-essentials-in-r/#leave-one-out-cross-validation---loocv>

Always cite your references

Dos & Don'ts of Technical Presentations

Dos

- Do tell a story with your slides.
- Do make sure key takeaway messages are clear.
- Do keep the number of slides to a minimum.
- Do ensure graphics are of good quality and are easy to interpret.
- Do make sure graphics aid your message (don't include figures for the sake of having something on your slide).
- Do know your audience and use an appropriate level of technical detail. Err on the side of less-technical.
- Do make your slides interesting!
- Do practice your presentation. This can be especially helpful if you don't like public speaking.

Don'ts

- Don't include too much information on one slide.
- Don't include too much information in one presentation.
- Don't fill your slides with text or equations.
- Don't include anything that you don't feel comfortable speaking about.