

Data is not flat

How can feature engineering help?

Alisa Dammer

me: alisdammer.com

@FedorinoGore_90

July 26, 2018

Structure

Feature engineering

Classification example

Conclusion

Questions?

What is feature engineering?

the process of using domain knowledge of the data to create features that make machine learning algorithms work

Feature engineering IS

1. Expert domain knowledge
2. Manual and/or automated process
3. Quite expensive to obtain
4. More relevant than the model

Feature engineering IS NOT

- ▶ Part of the preprocessing
- ▶ Part of the initial data collection

Data Engineers are not responsible for this step

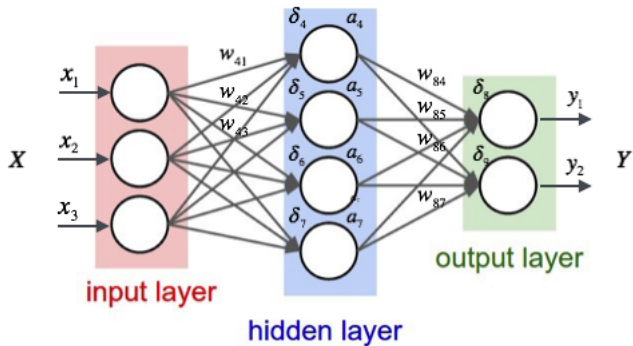
Approaches

1. automated feature engineering
2. manual feature engineering

Where to begin

- ▶ Brainstorming or Testing features
- ▶ Deciding what features to create
- ▶ Creating features
- ▶ Checking how the features work with your model
- ▶ Improving your features if needed
- ▶ Go back to brainstorming/creating more features until the work is done

Example



http://localhost:8888/notebooks/classification_example.ipynb

So, what can I say?

- ▶ Costly
- ▶ Error prone
- ▶ FE is not a knowledge transfer
- ▶ Enables variety of methods (Features selection, Automated Feature creation, etc)
- ▶ Very powerful
- ▶ Works with all models (be smart about it tho)

Thanks for the patience!



Maybe useful information

- ▶ https://en.wikipedia.org/wiki/Feature_engineering
- ▶ <https://towardsdatascience.com/understanding-feature-engineering-part-1-continuous-numeric-data-da4e47099a71>
- ▶ "Applied Predictive Modeling" Max Kuhn, Kjell Johson