Feature engineering process

Here we describe a basic feature engineering process. We'll cover more of the individual part of this process as the course progresses.

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 complete

What is Feature Engineering?

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

You can see the dependencies in this definition:

The performance measures you’ve chosen (RMSE? AUC?)

The framing of the problem (classification? regression?)

The predictive models you’re using (SVM?)

The raw data you have selected and prepared (samples? formatting? cleaning?)

Process of Machine Learning

(tasks before here…)

Select Data: Integrate data, de-normalize it into a dataset, collect it together.

Preprocess Data: Format it, clean it, sample it so you can work with it.

Transform Data: Feature Engineer happens here.

Model Data: Create models, evaluate them and tune them.

(tasks after here…)

Iterative Process of Feature Engineering

Brainstorm features: Really get into the problem, look at a lot of data, study feature engineering on other problems and see what you can steal.

Devise features: Depends on your problem, but you may use automatic feature extraction, manual feature construction and mixtures of the two.

Select features: Use different feature importance scorings and feature selection methods to prepare one or more “views” for your models to operate upon.

Evaluate models: Estimate model accuracy on unseen data using the chosen features.