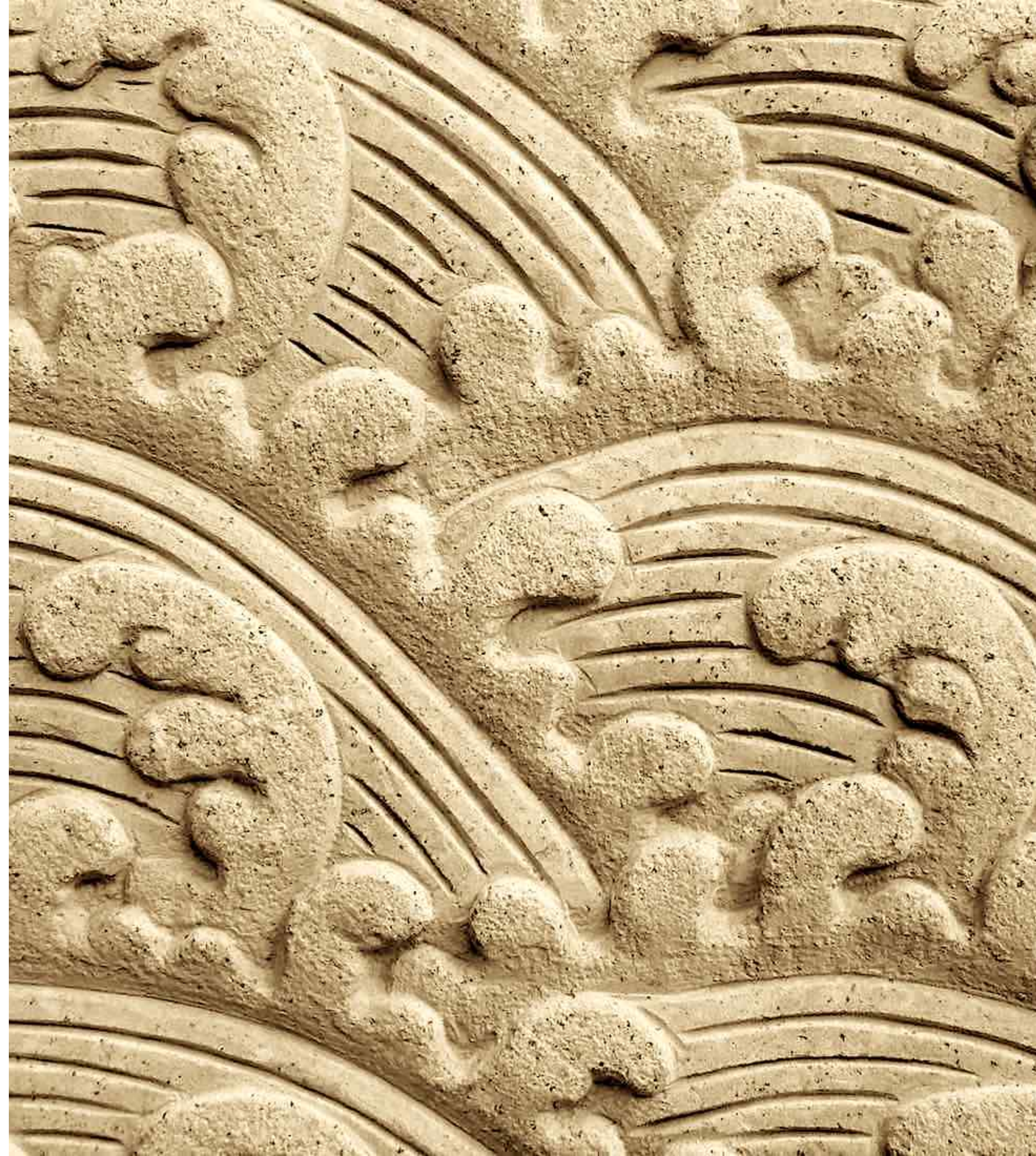


Speech Processing

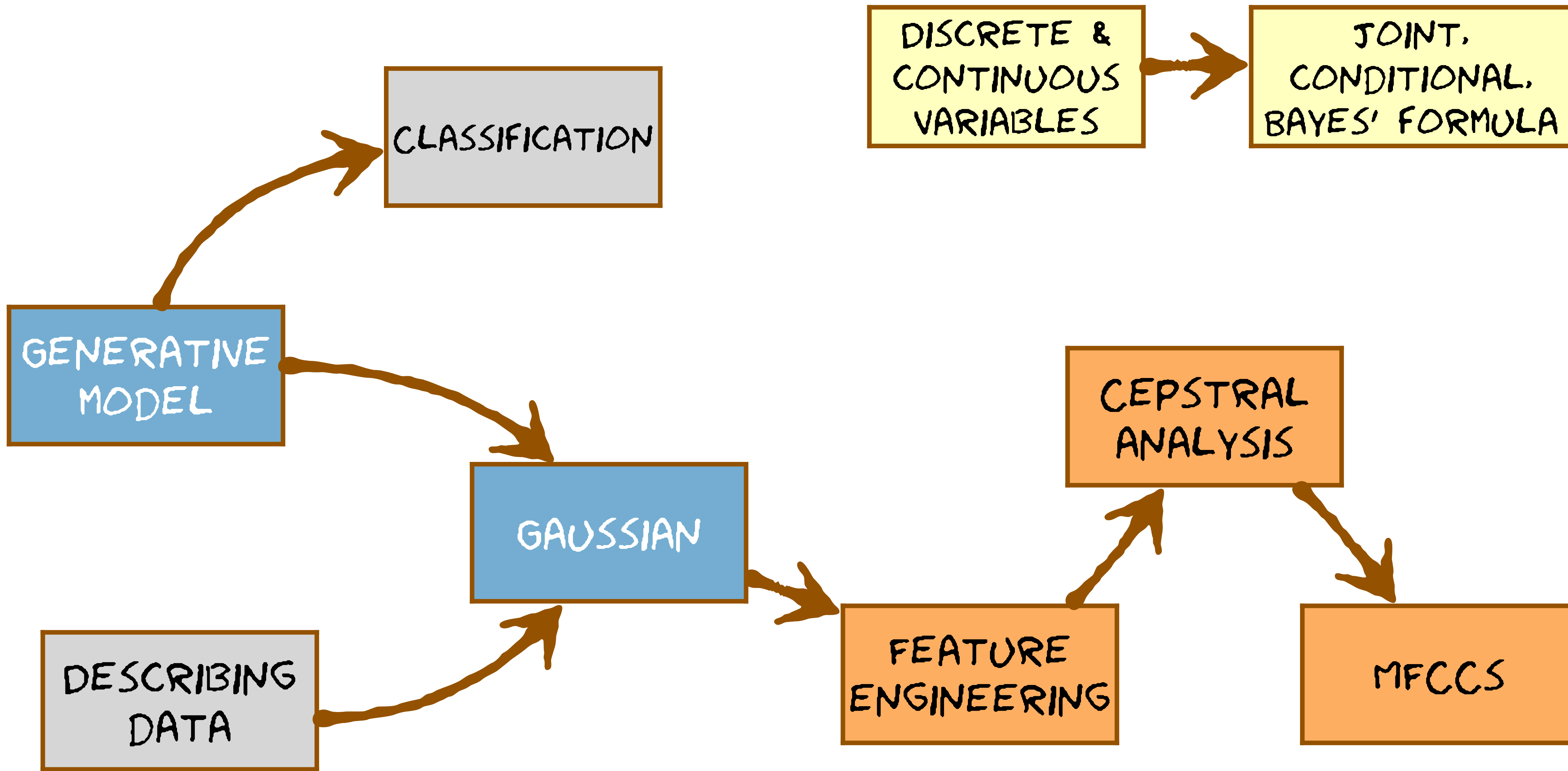
Simon King
University of Edinburgh

additional class slides for 2020-21



Module 7

Feature engineering



Orientation

- We're on a journey towards HMMs
- Pattern **matching**
- Extracting **features** from speech
- Probabilistic **generative** modelling

What we are learning along the way



Dynamic programming
(in the form of Dynamic Time Warping)

The interaction between

- choice of model
- choice of features

Dynamic programming
(in the form of the Viterbi algorithm)

What you should already know

- Probability

- the Gaussian probability density function
- **covariance**, and why we'd prefer **not** to have to model it

- Human hearing

- non-linear frequency resolution
- amplitude compression
- the cochlea is like a filterbank

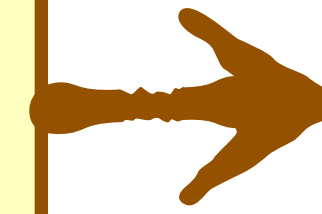
Massively increases the number of parameters. That would require a lot more training data.

Useful **inspiration** for feature extraction

GENERATIVE
MODEL

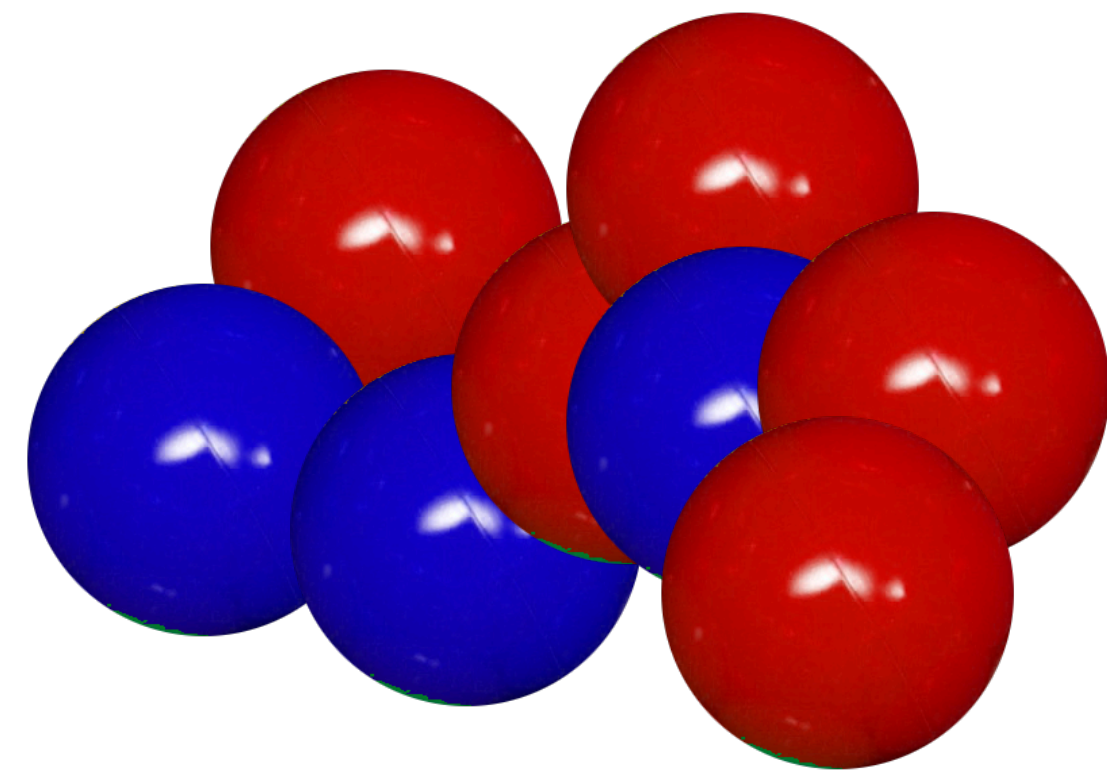
DESCRIBING
DATA

DISCRETE &
CONTINUOUS
VARIABLES



JOINT,
CONDITIONAL,
BAYES' FORMULA

A conceptual leap: generative models



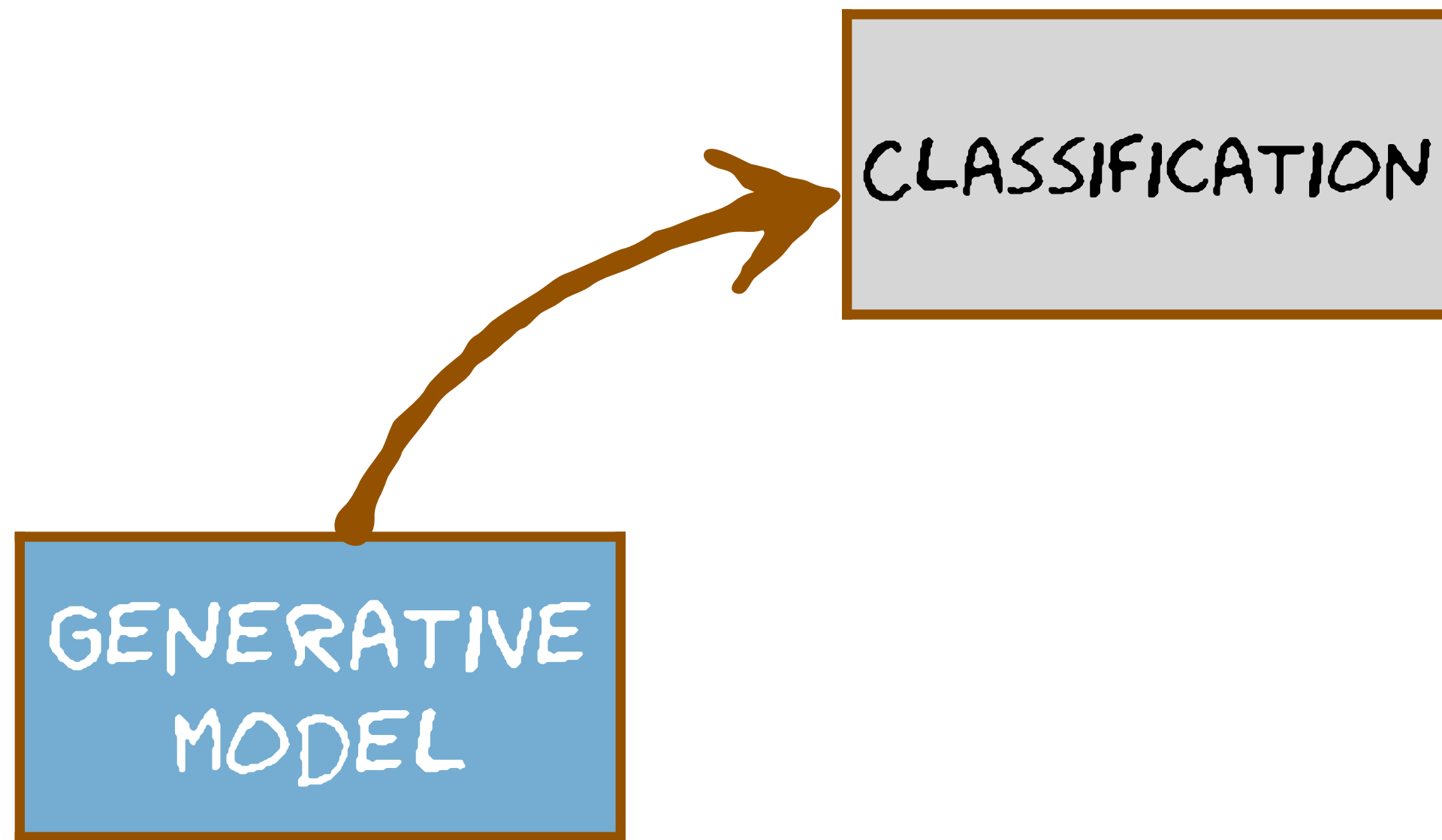
Model A



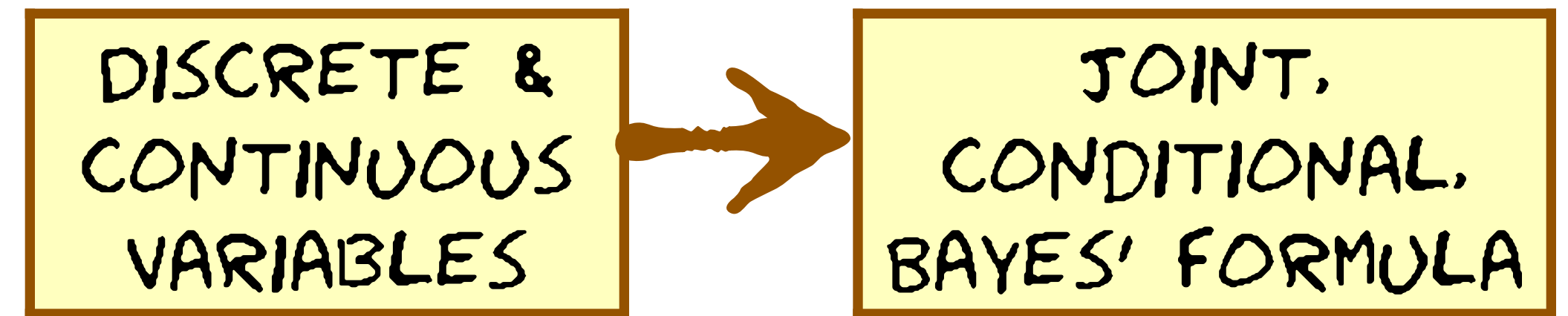
Model B



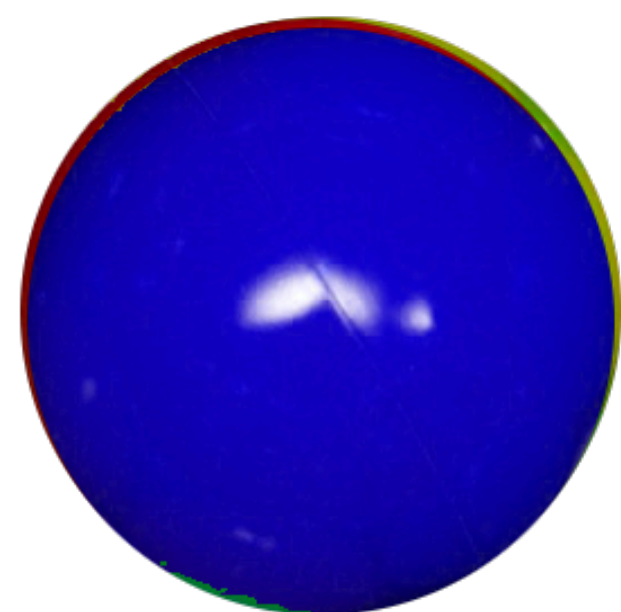
Model C



DESCRIBING
DATA



Generative models performing classification



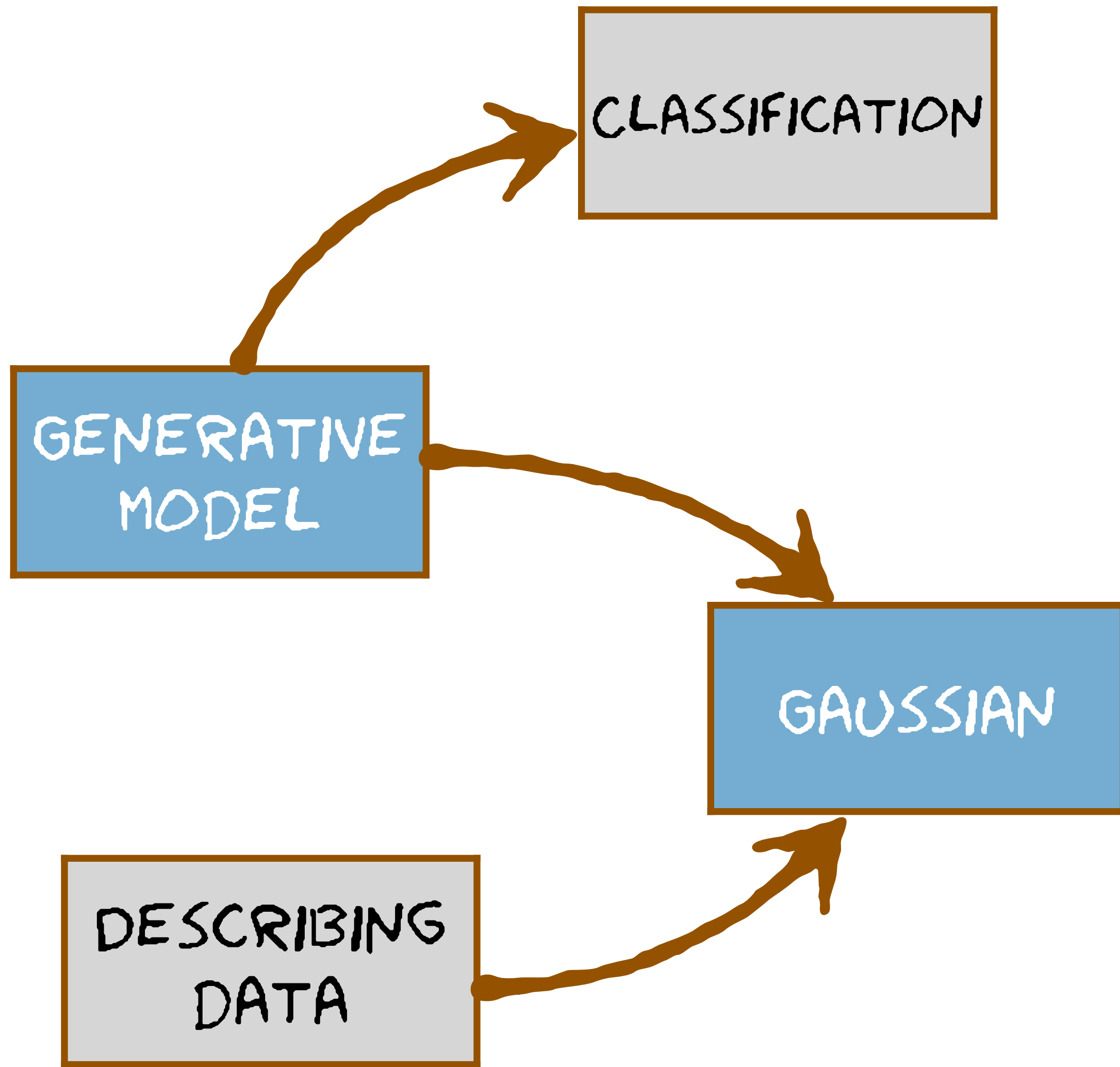
Model A



Model B



Model C



Describing data with the Gaussian probability density function

The Gaussian as a generative model



Model A

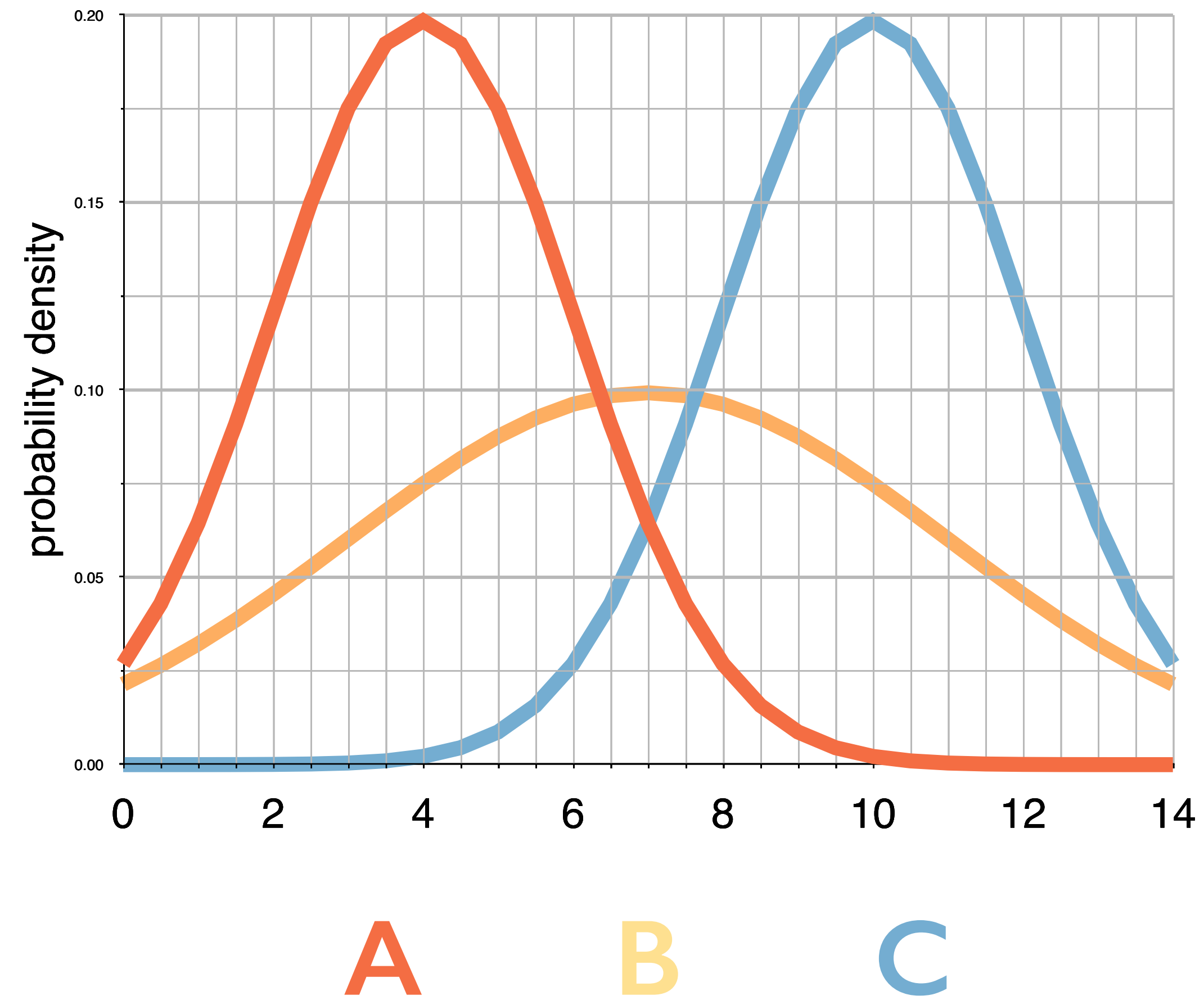


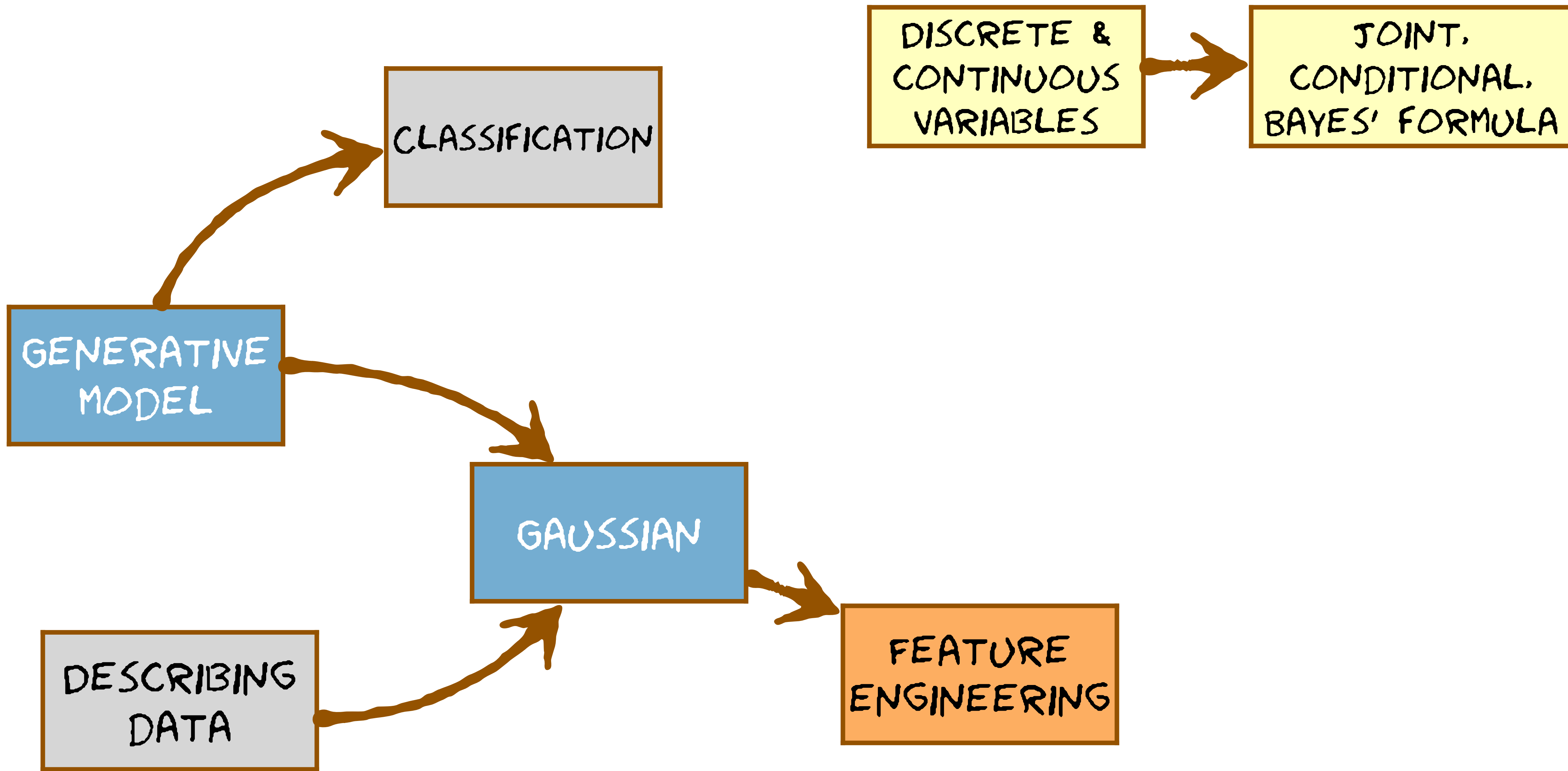
Model B



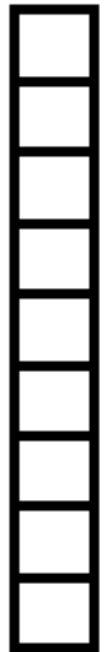
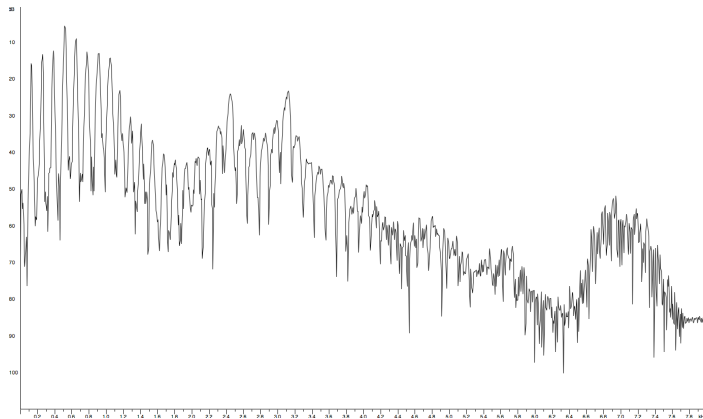
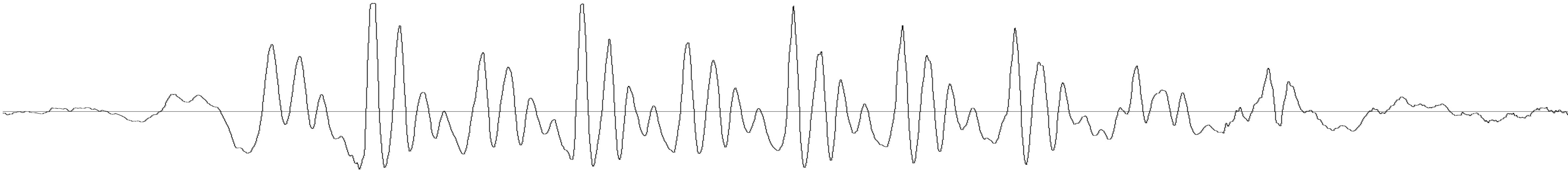
Model C

Gaussian generative models performing classification

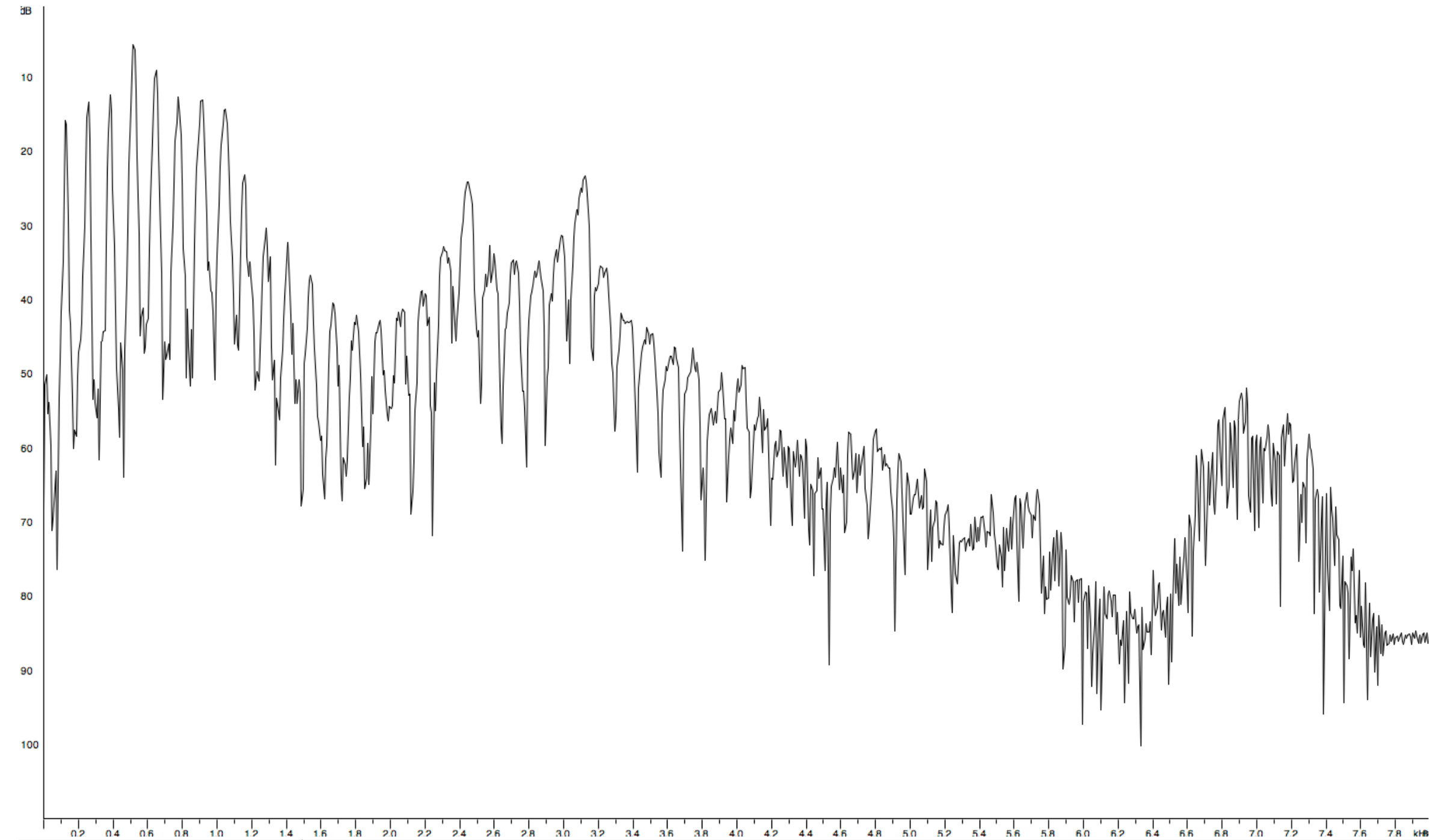




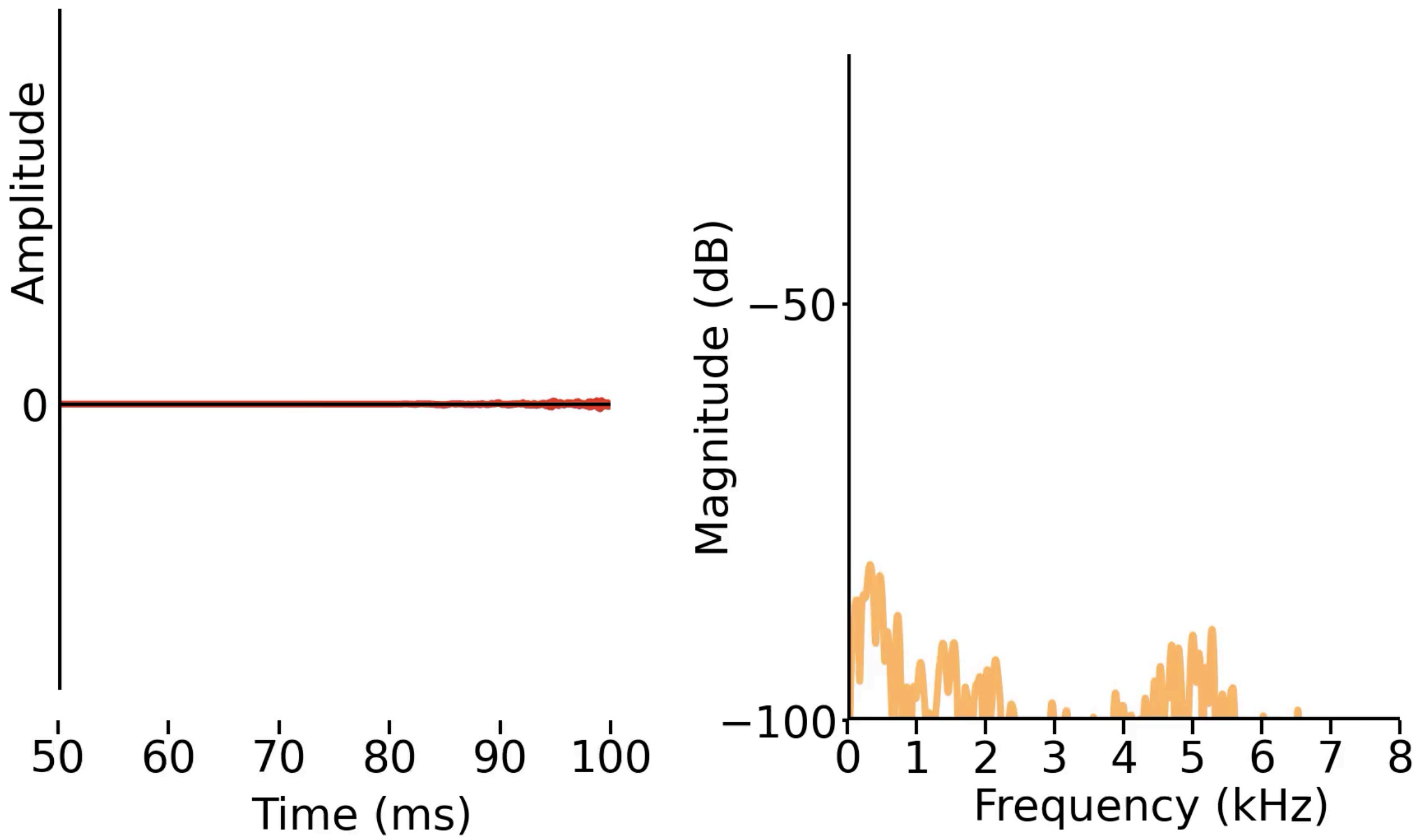
Recap: Filterbank features for automatic speech recognition

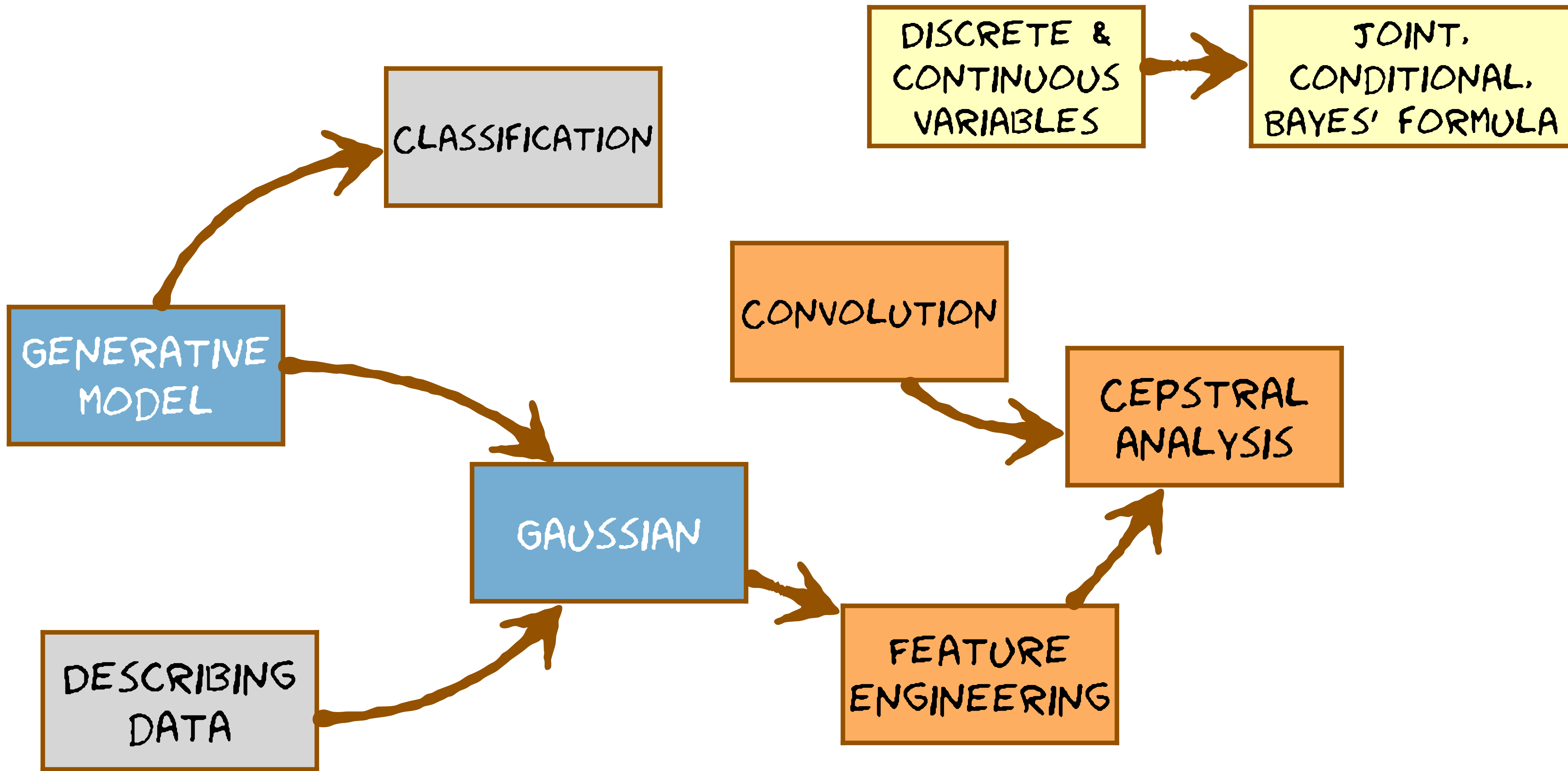


Recap: Filterbank features for automatic speech recognition

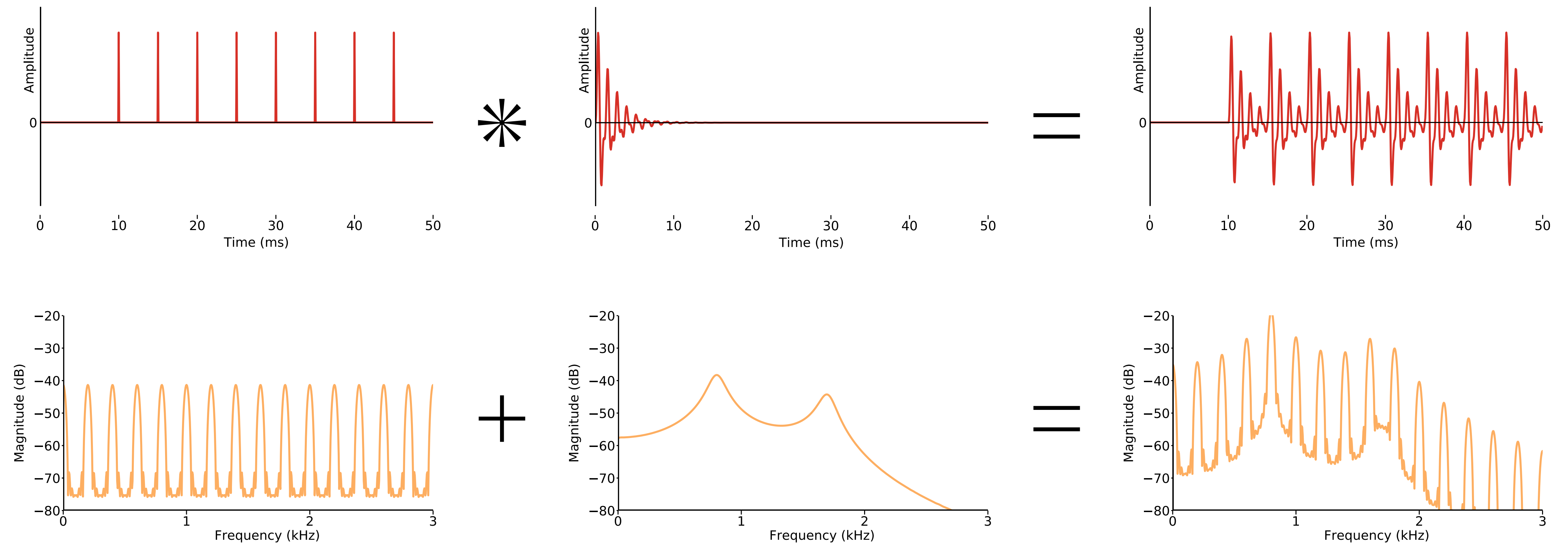


Correlation between features \Rightarrow need to model covariance

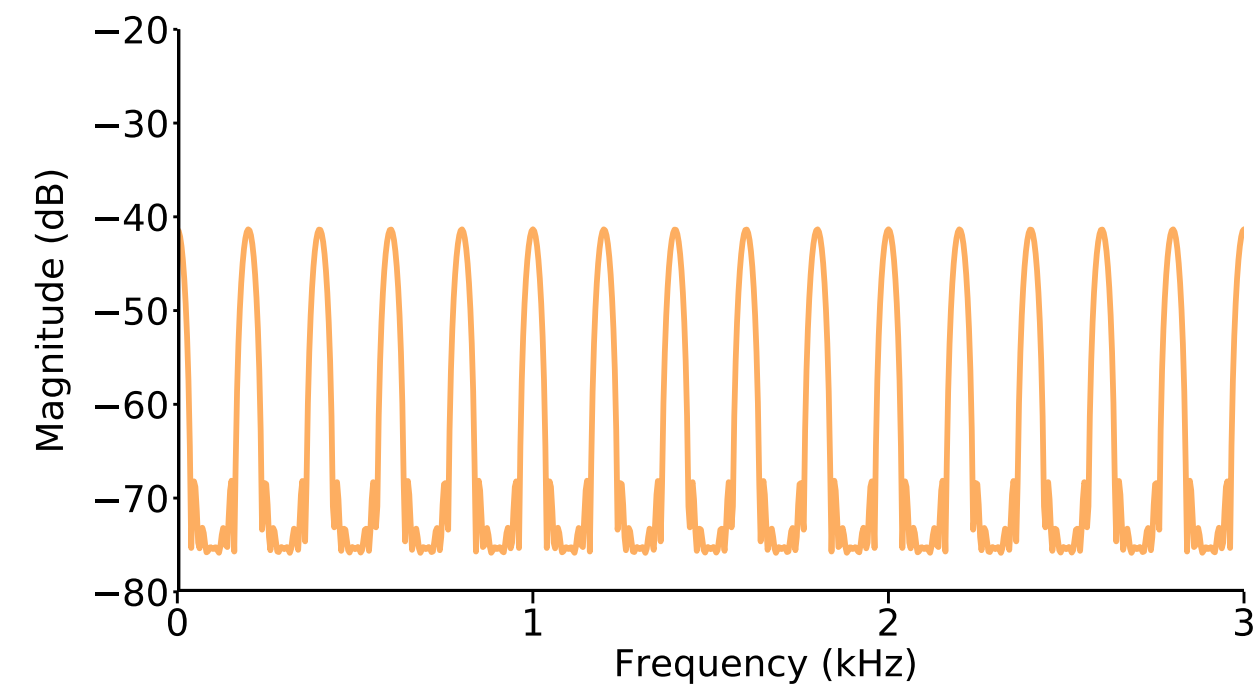




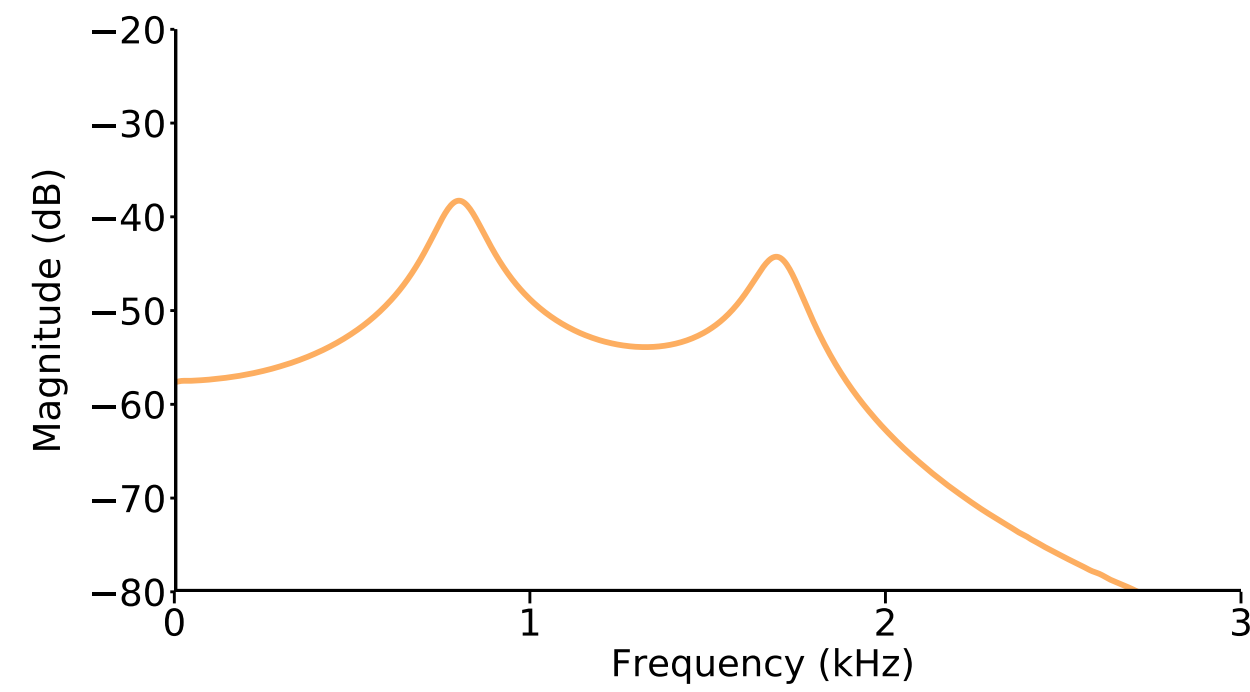
recap: convolution of waveforms = addition of log magnitude spectra



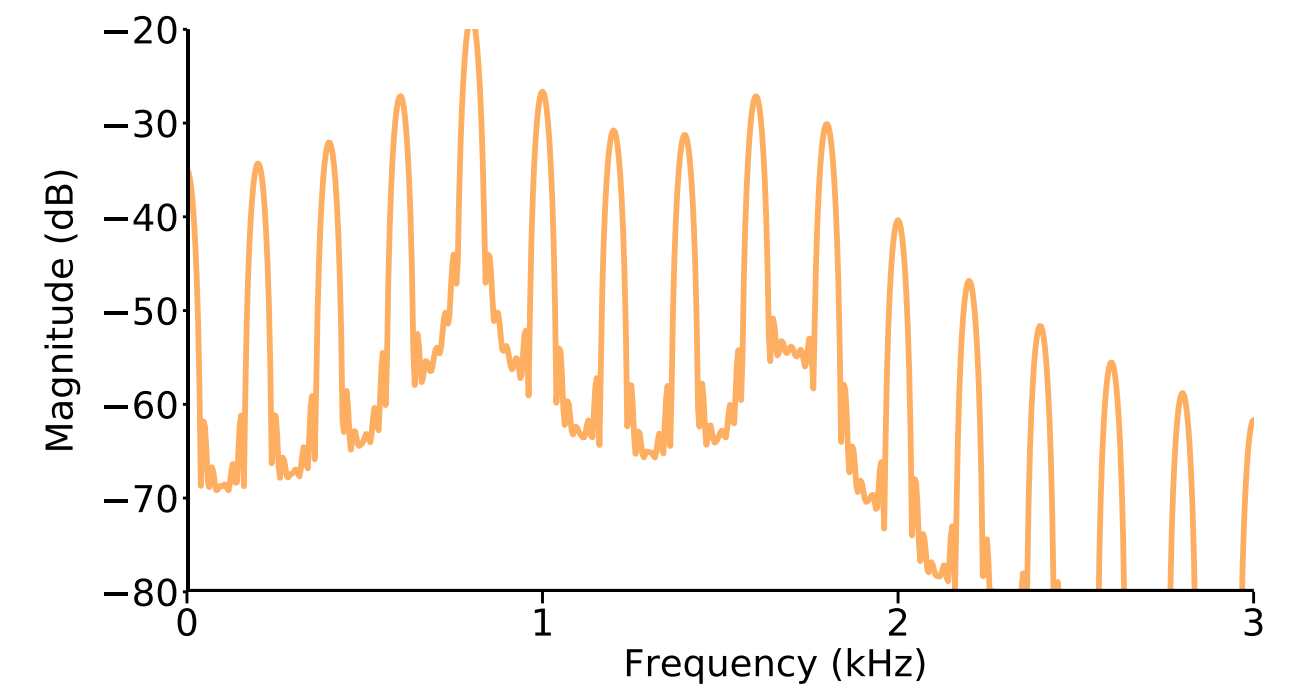
addition of log magnitude spectra



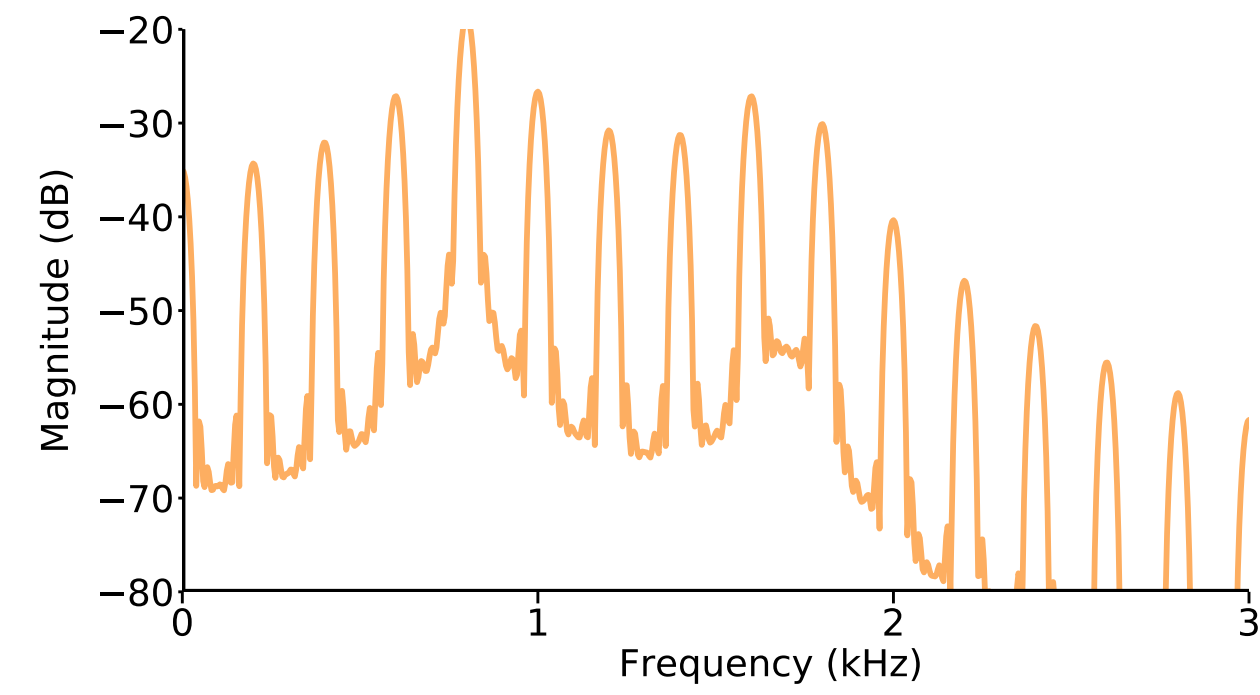
+



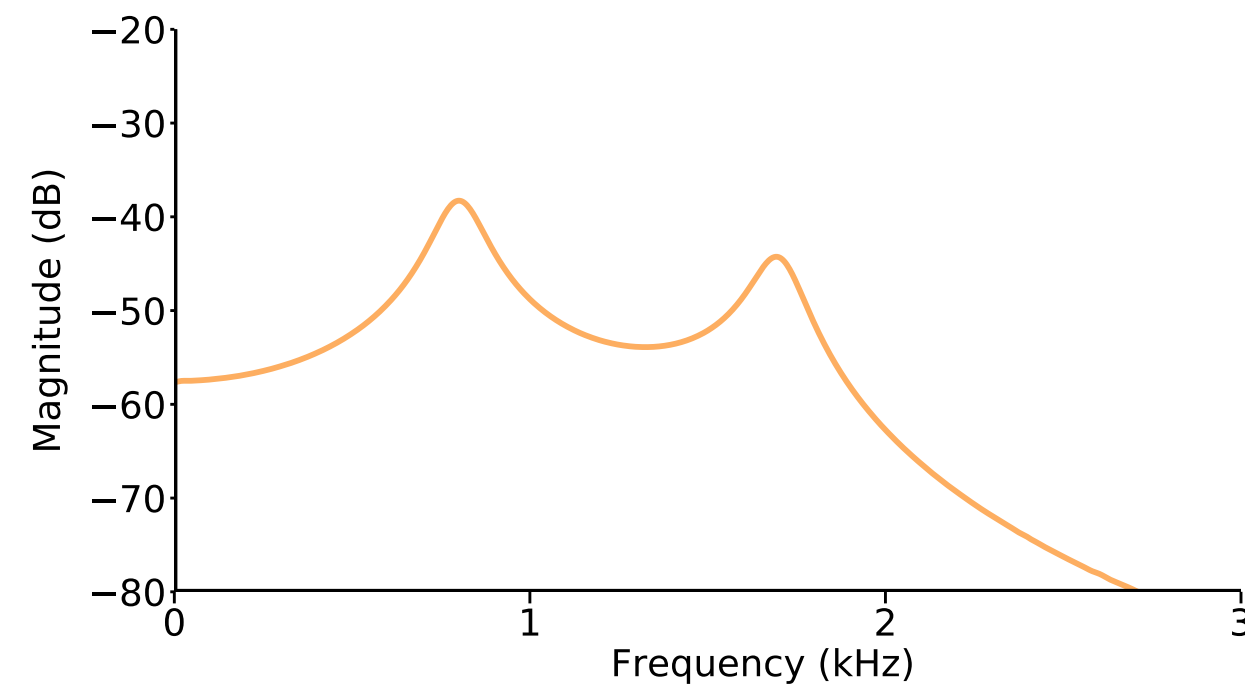
=



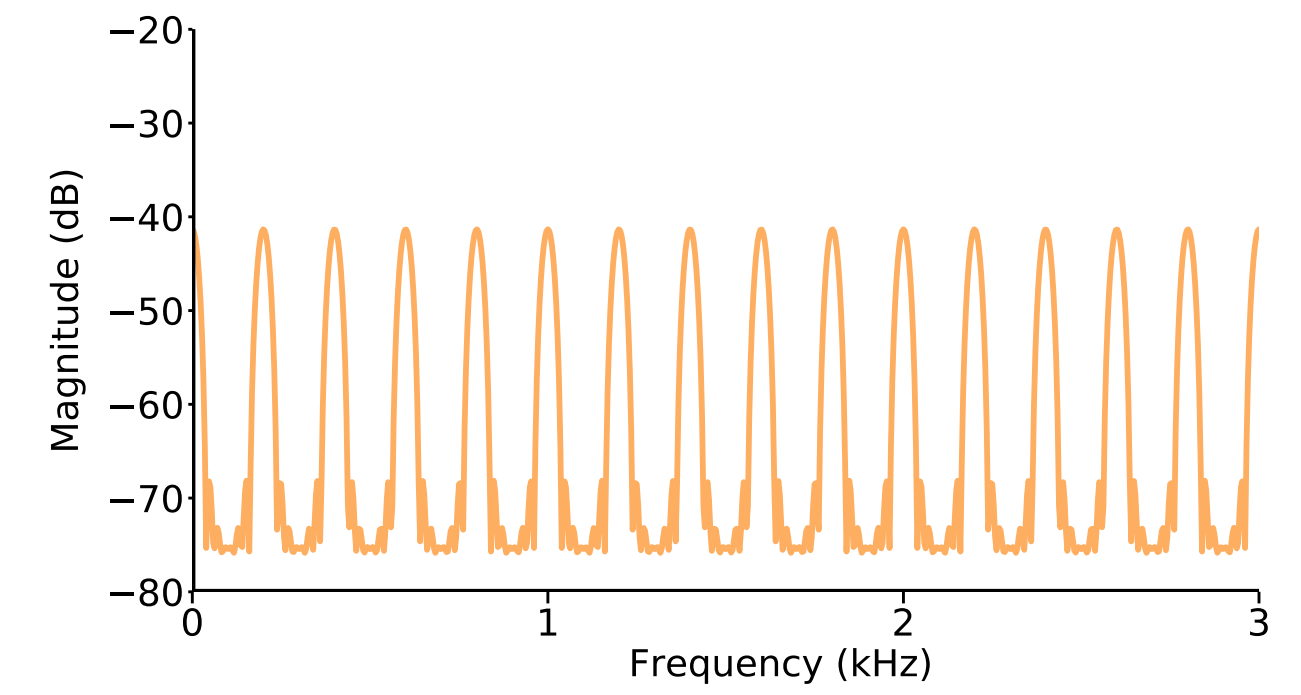
but we want to do this



=

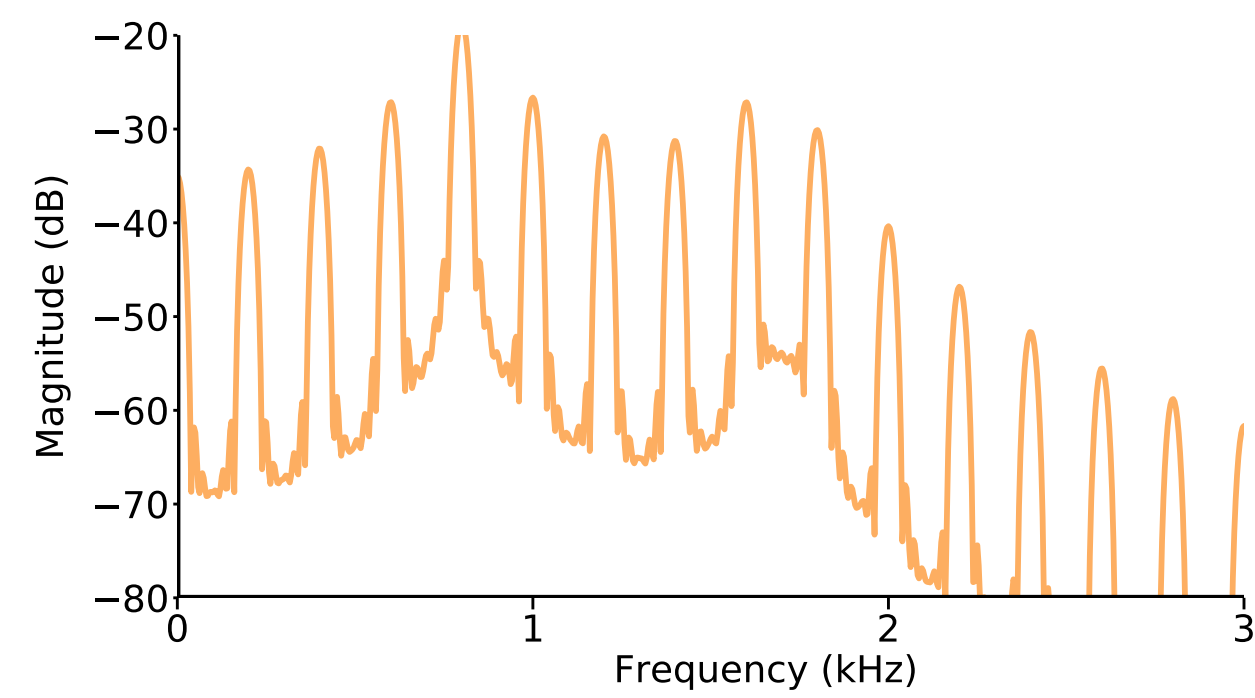


+

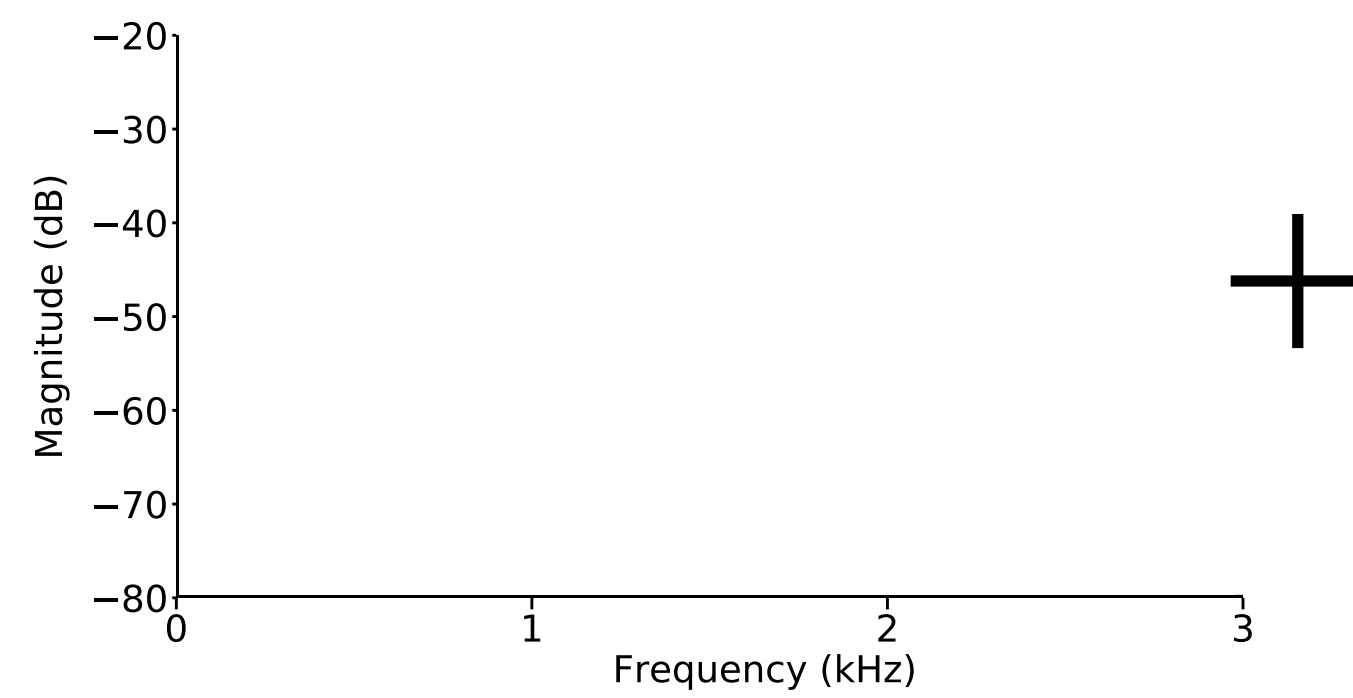


SERIES EXPANSION

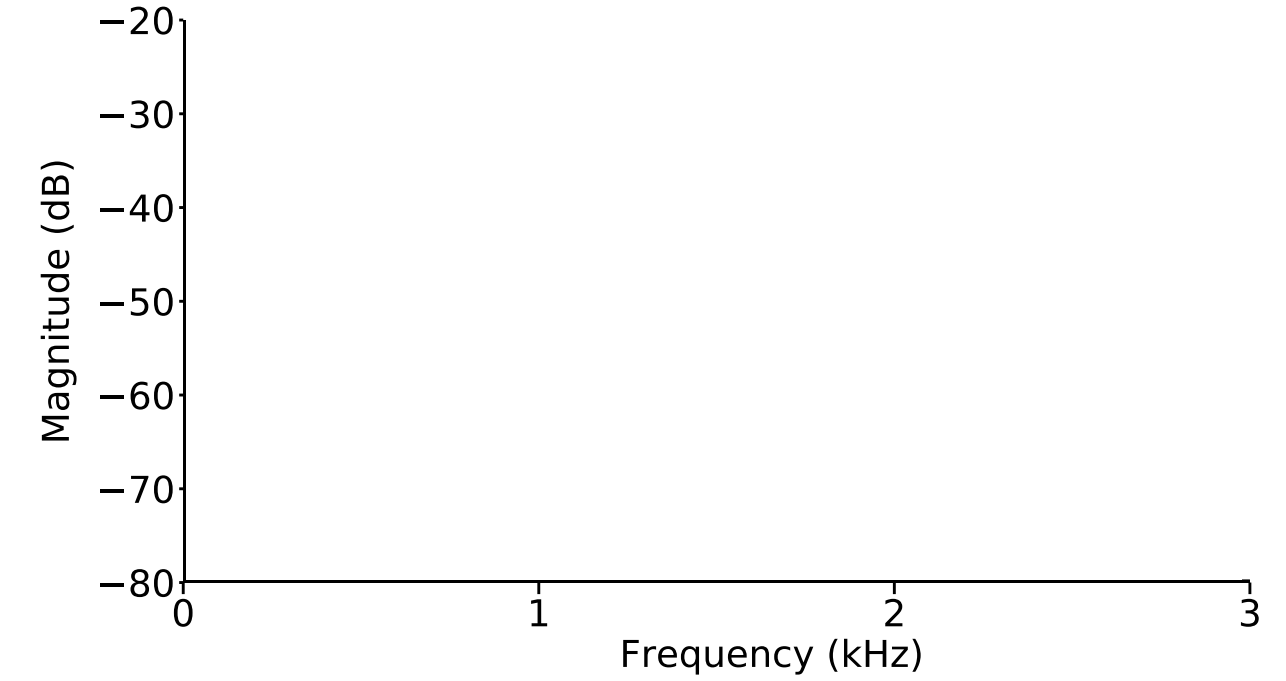
a more general expression that we can solve



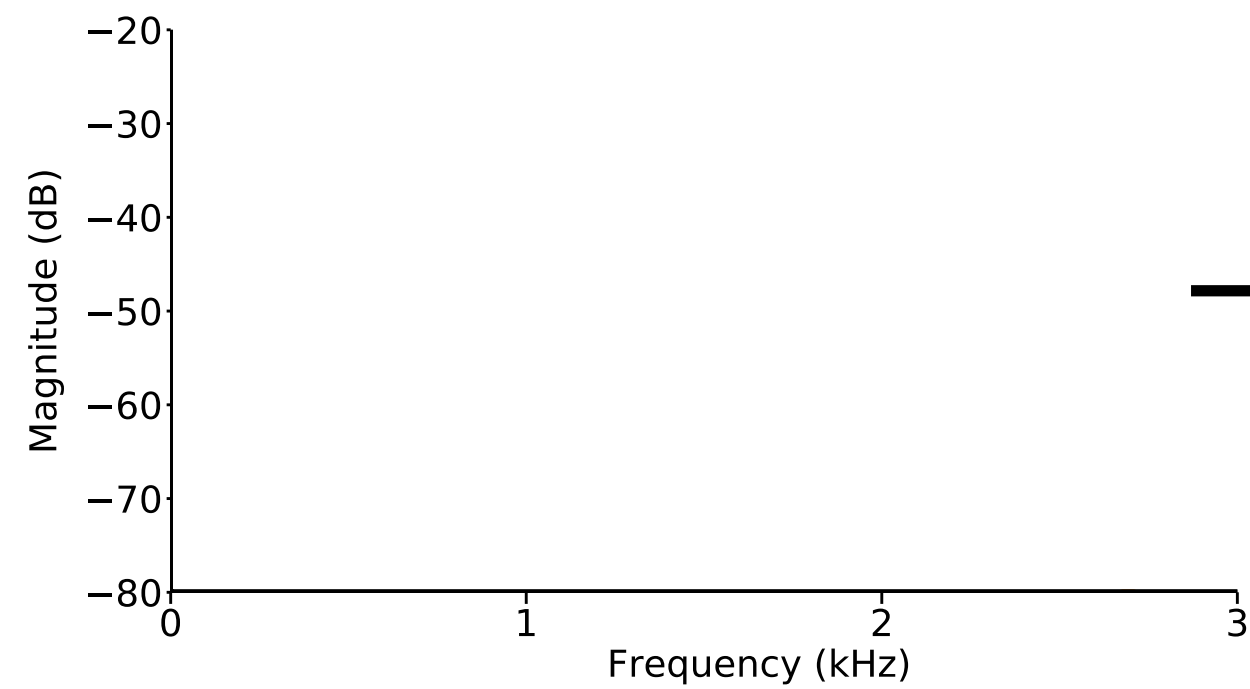
=



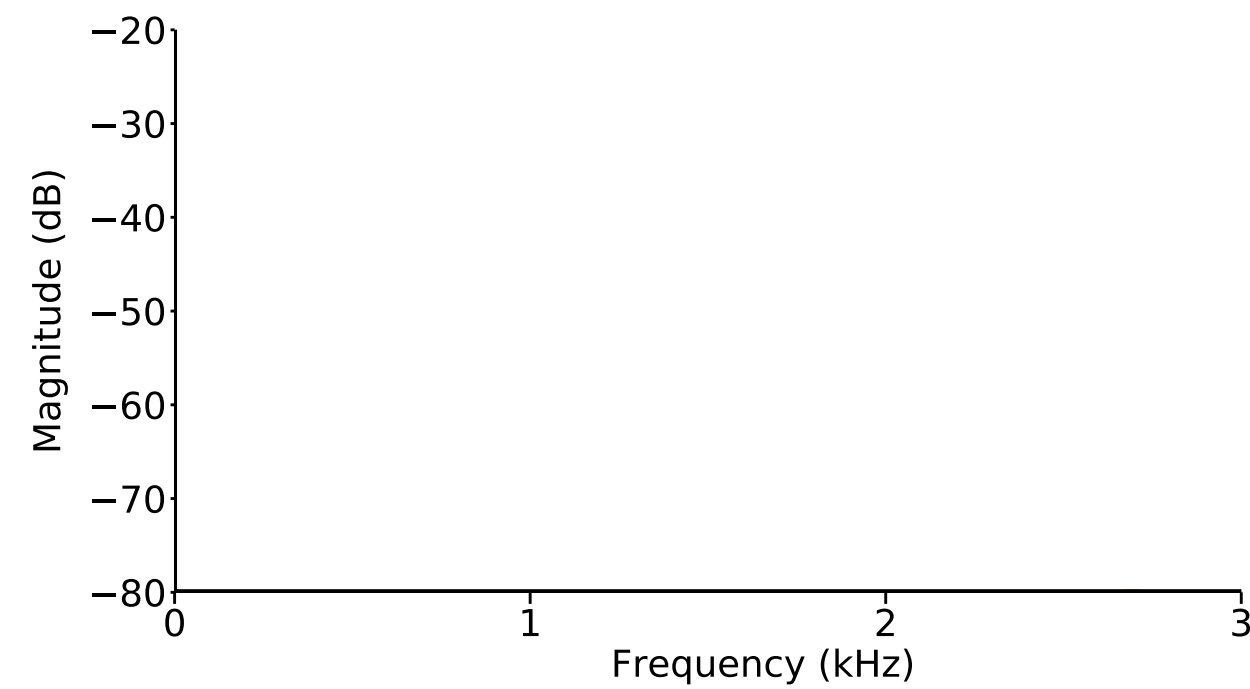
+



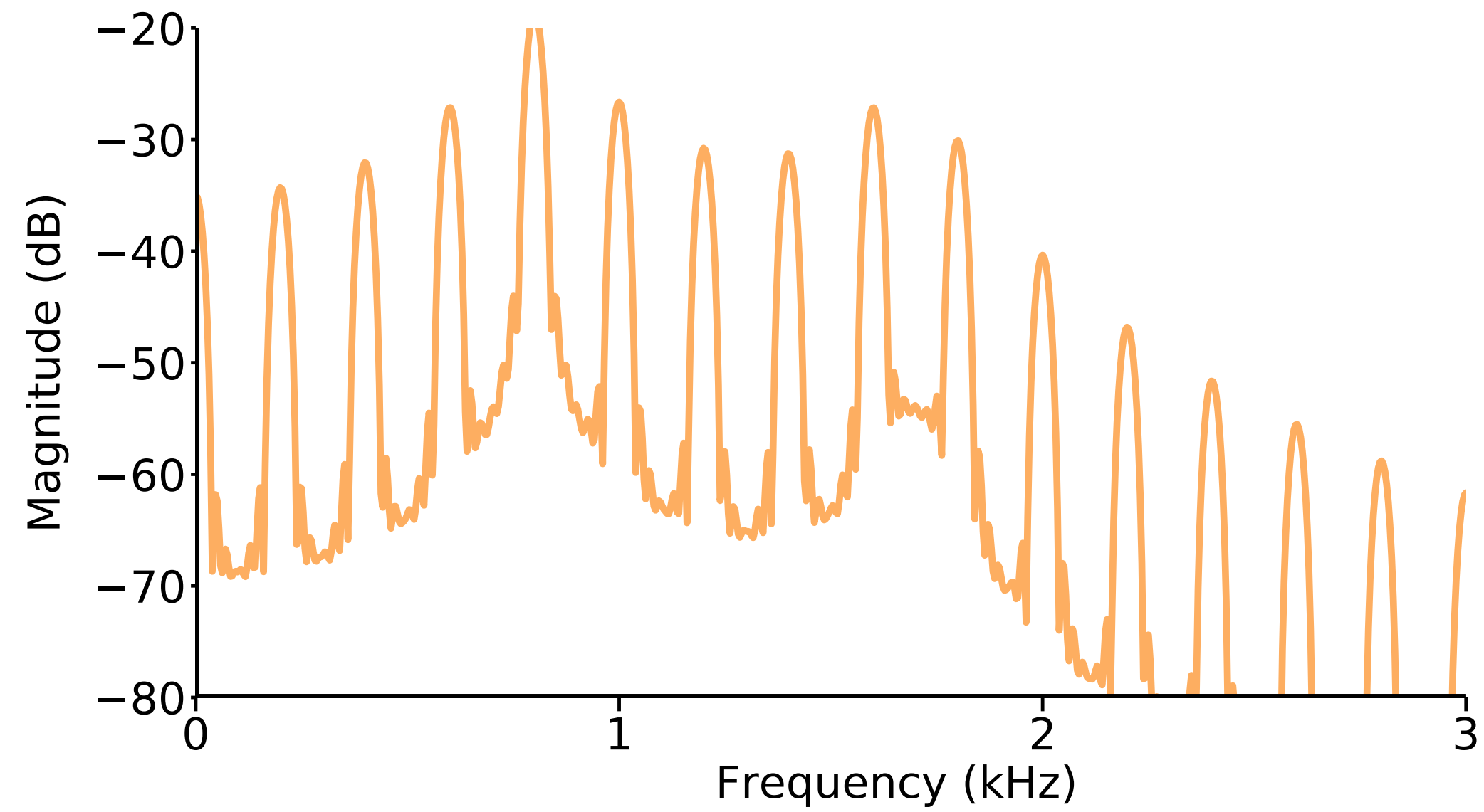
+



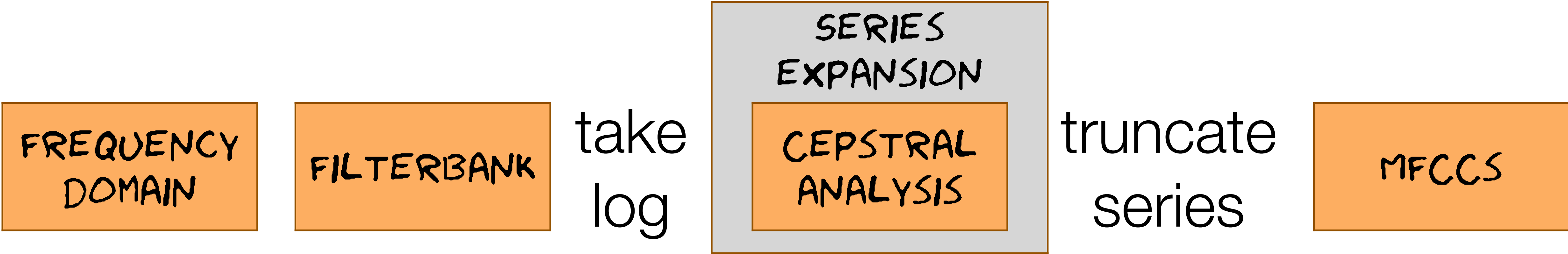
+



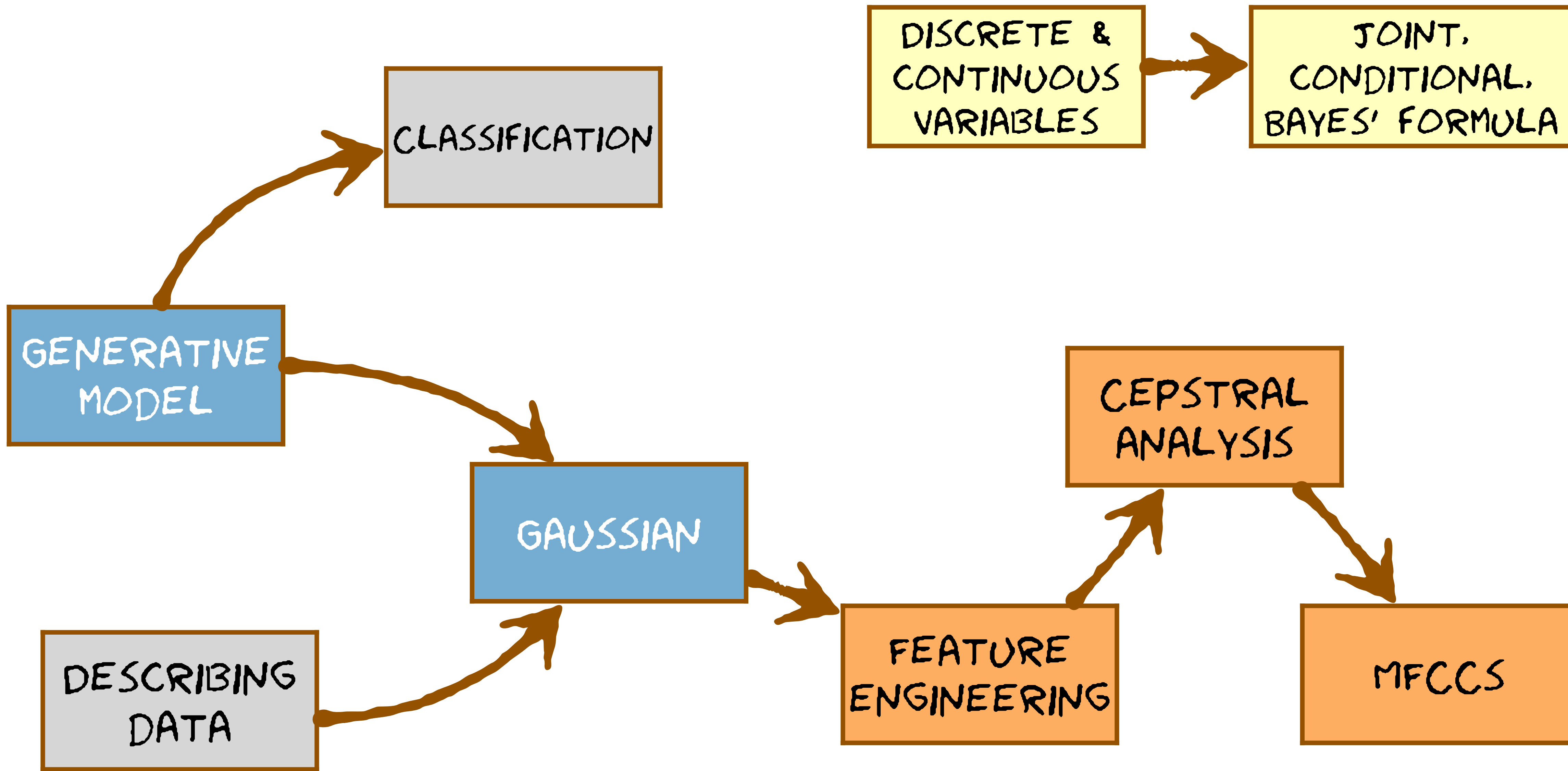
The spectrum and the cepstrum



Feature engineering: Mel Frequency Cepstral Coefficients



But did we remove covariance?



What next?

- From the Gaussian generative model to a model that **generates a sequence**
 - the Hidden Markov Model (HMM)
- Deciding what to model
 - whole words ?
 - sub-word units ?
- Connected speech
- Estimating the parameters of the HMM

Module 8

Module 9