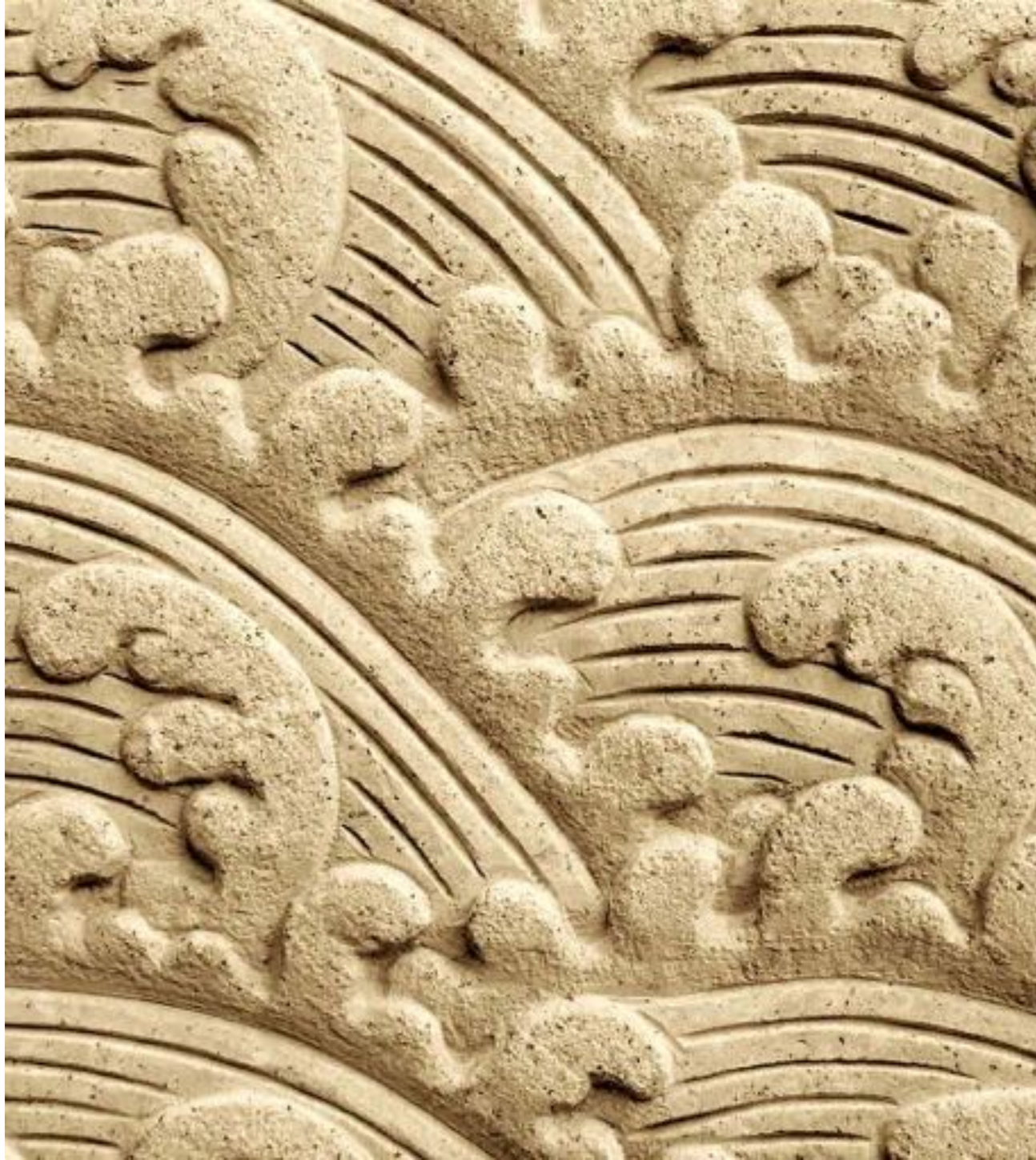
Speech Processing

Simon King
University of Edinburgh

2022-23



Module 8

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)

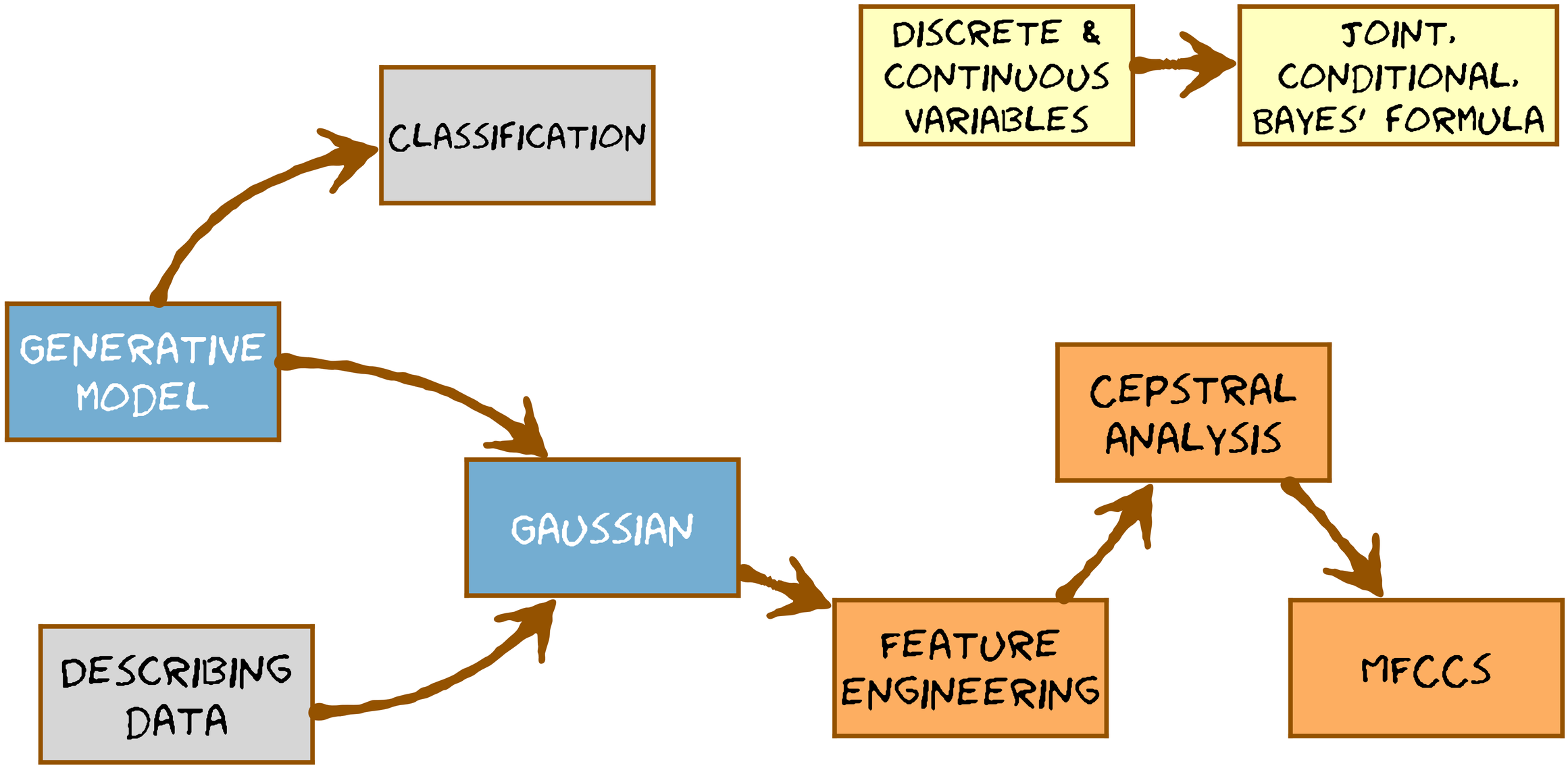
What you should already know

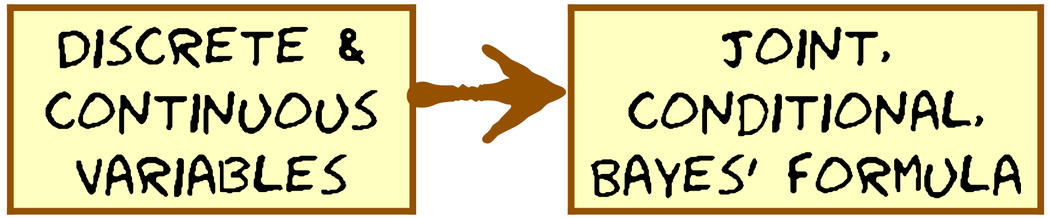
- Probability
  - the Gaussian probability density function
  - **covariance**, and why we'd prefer **not** to have to model it

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

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

Useful **inspiration** for feature extraction

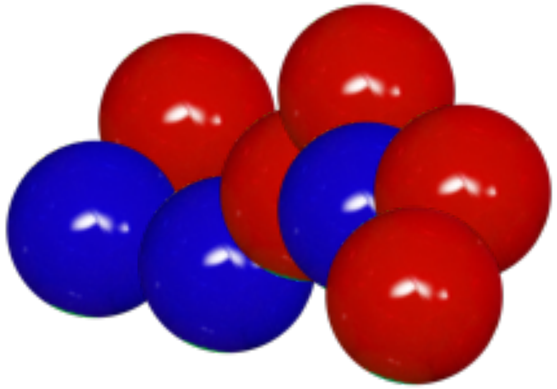

DISCRETE & CONTINUOUS VARIABLES
JOINT, CONDITIONAL, BAYES' FORMULA
CLASSIFICATION
GENERATIVE MODEL
GAUSSIAN
DESCRIBING DATA
CEPSTRAL ANALYSIS
FEATURE ENGINEERING
MFCCS



GENERATIVE MODEL

DESCRIBING DATA

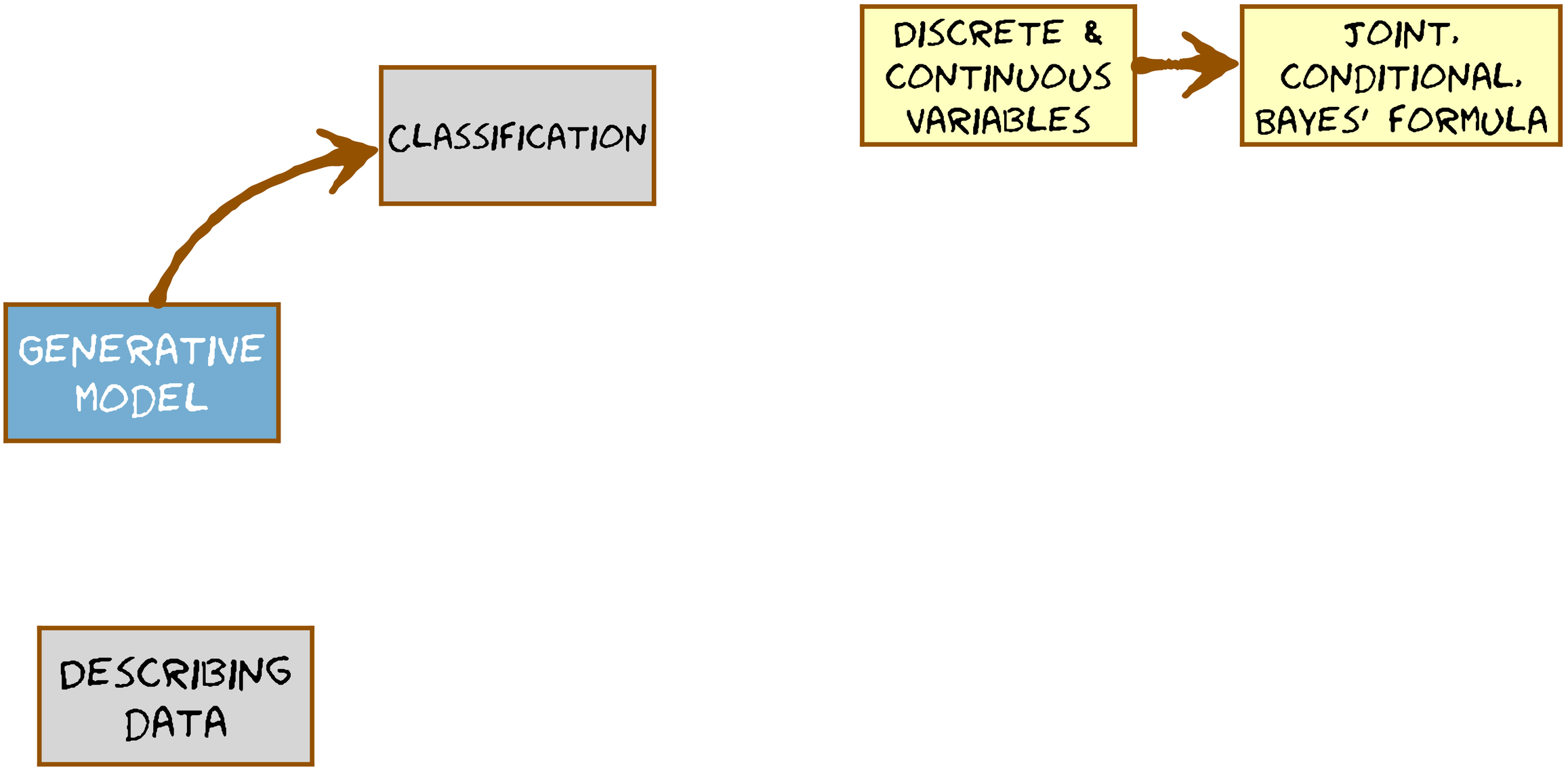
A conceptual leap: generative models



Model A

Model B

Model C

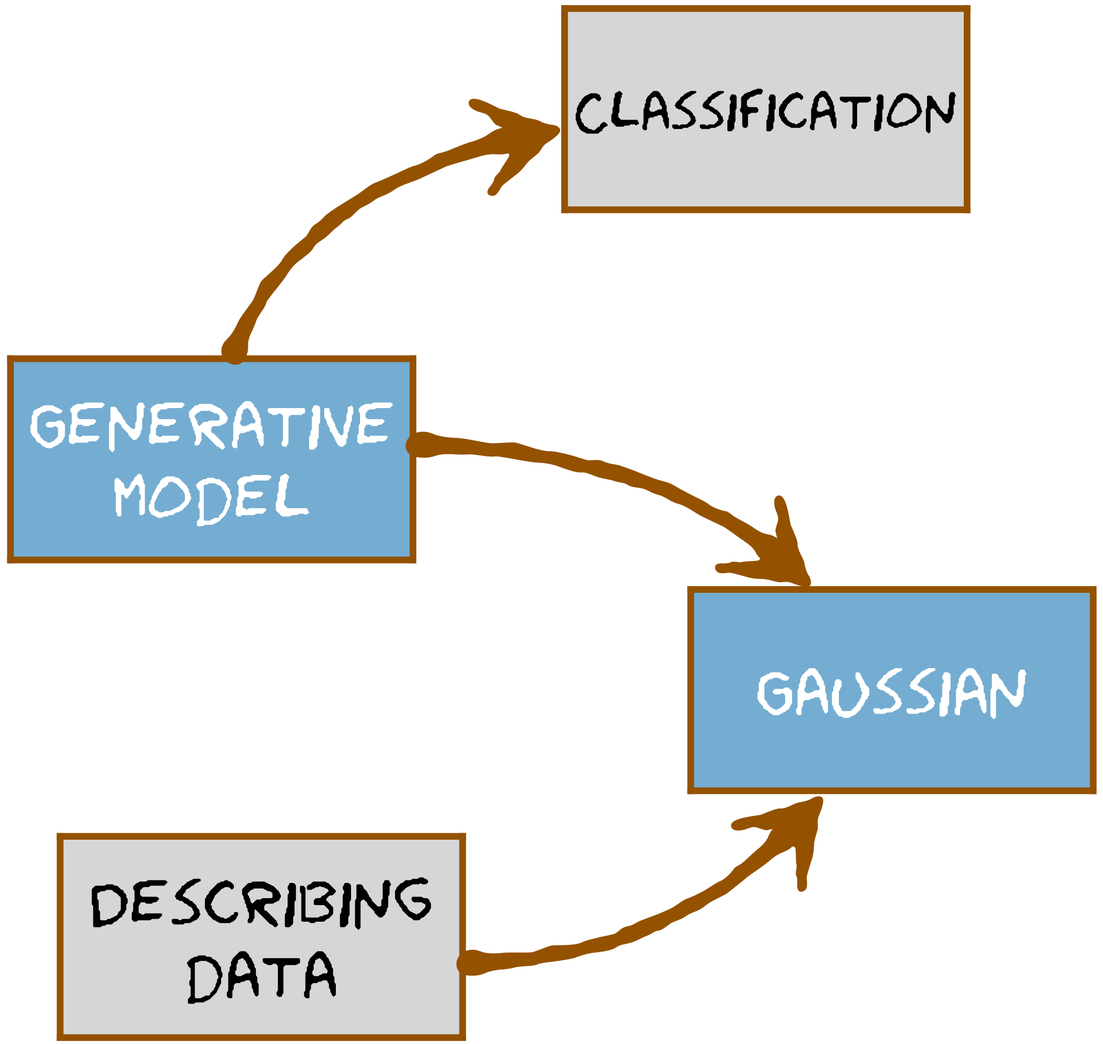

CLASSIFICATION
GENERATIVE MODEL
DISCRETE & CONTINUOUS VARIABLES
JOINT, CONDITIONAL, BAYES' FORMULA
DESCRIBING DATA

Generative models performing classification

Model A

Model B

Model C


CLASSIFICATION
GENERATIVE MODEL
GAUSSIAN
DESCRIBING DATA

Describing data with the Gaussian probability density function

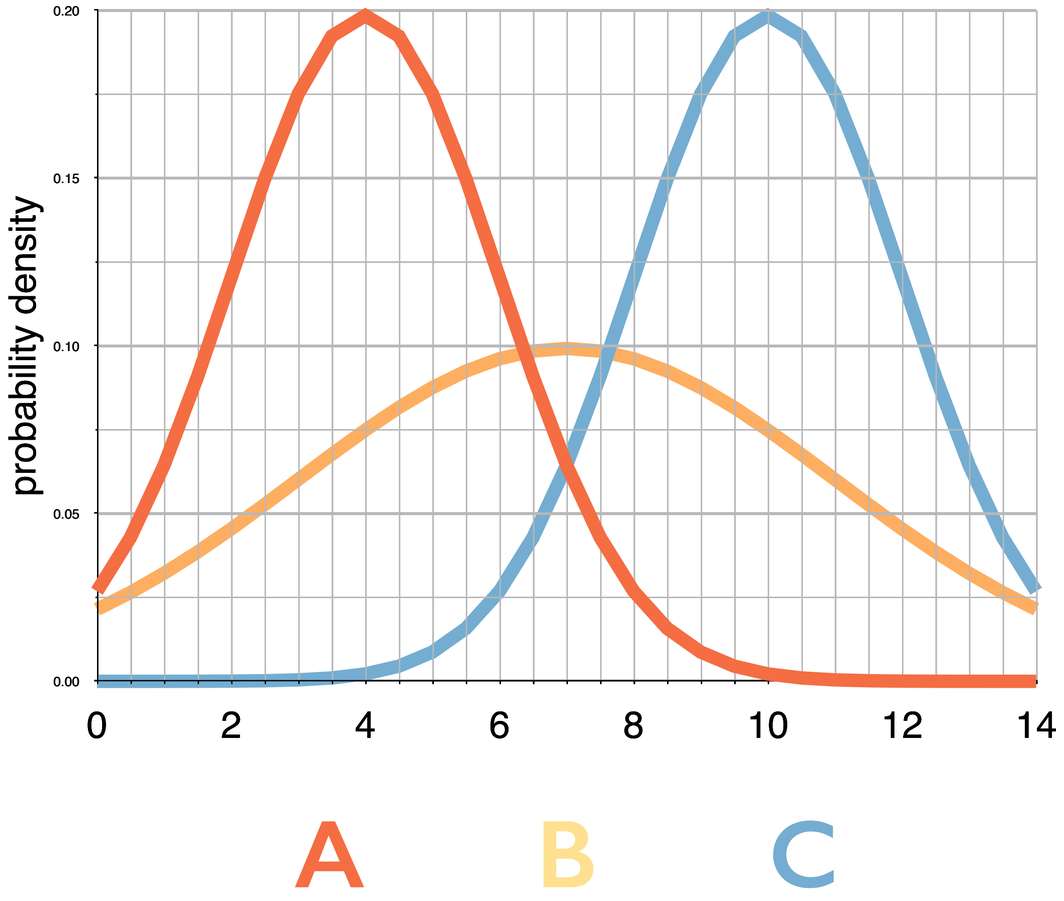
The Gaussian as a generative model

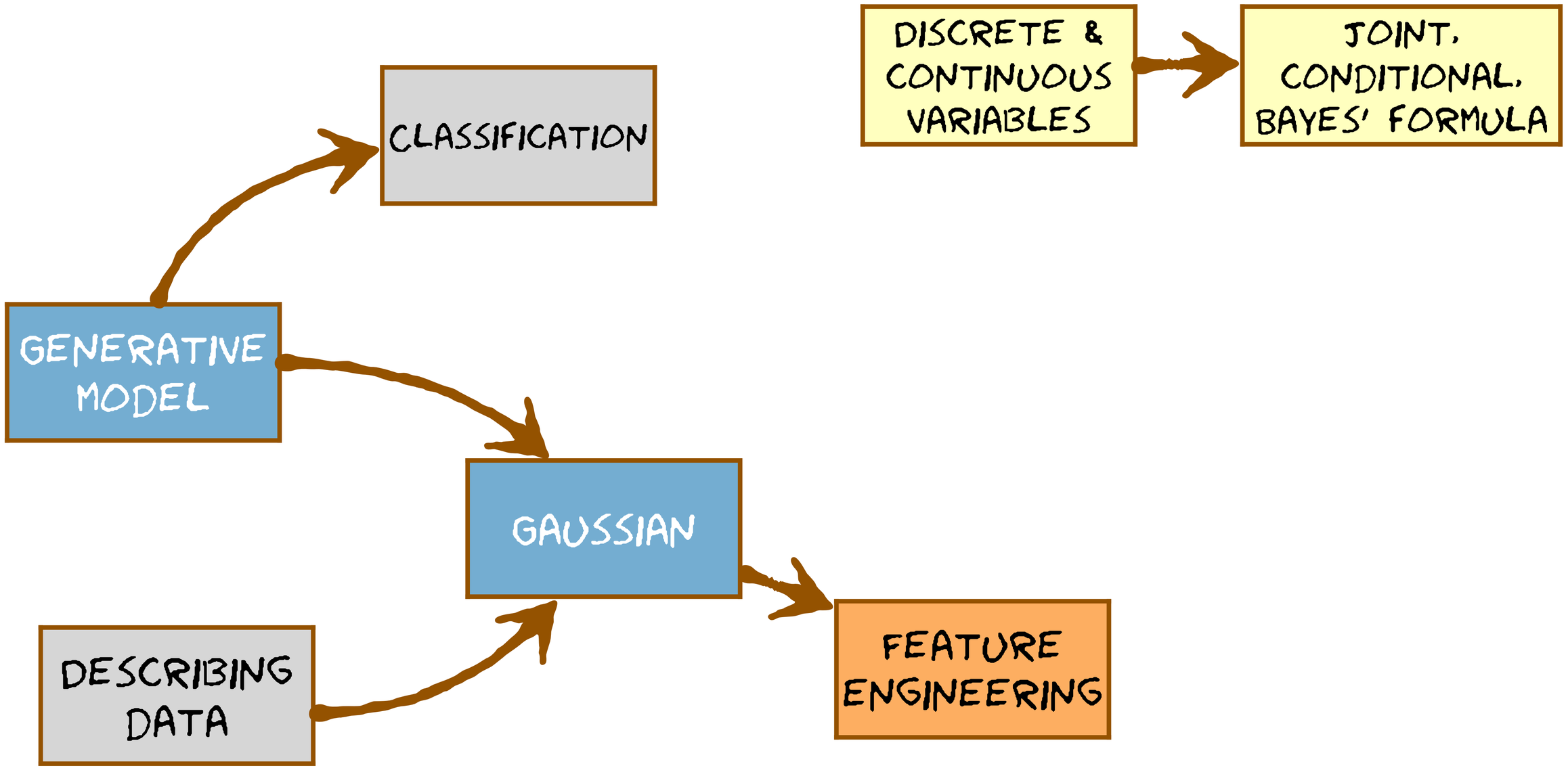
Model A

Model B

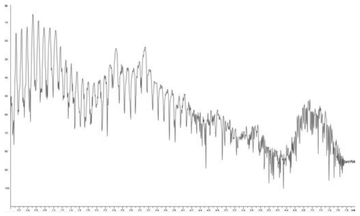
Model C

Gaussian generative models performing classification

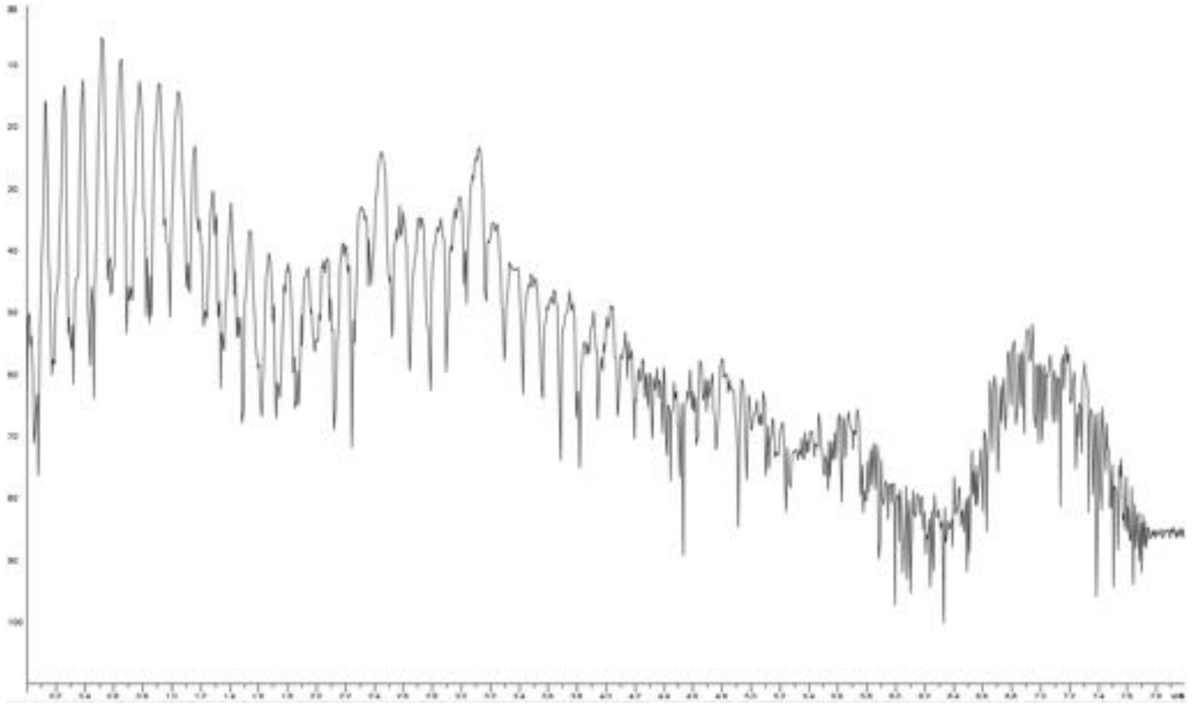


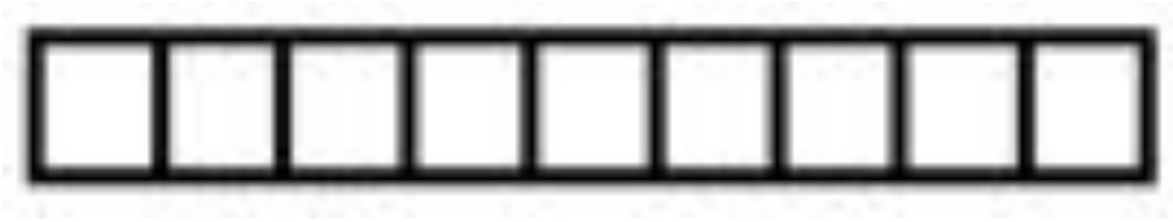

DISCRETE & CONTINUOUS VARIABLES
JOINT, CONDITIONAL, BAYES' FORMULA
CLASSIFICATION
GENERATIVE MODEL
GAUSSIAN
DESCRIBING DATA
FEATURE ENGINEERING

Recap: Filterbank features for automatic speech recognition

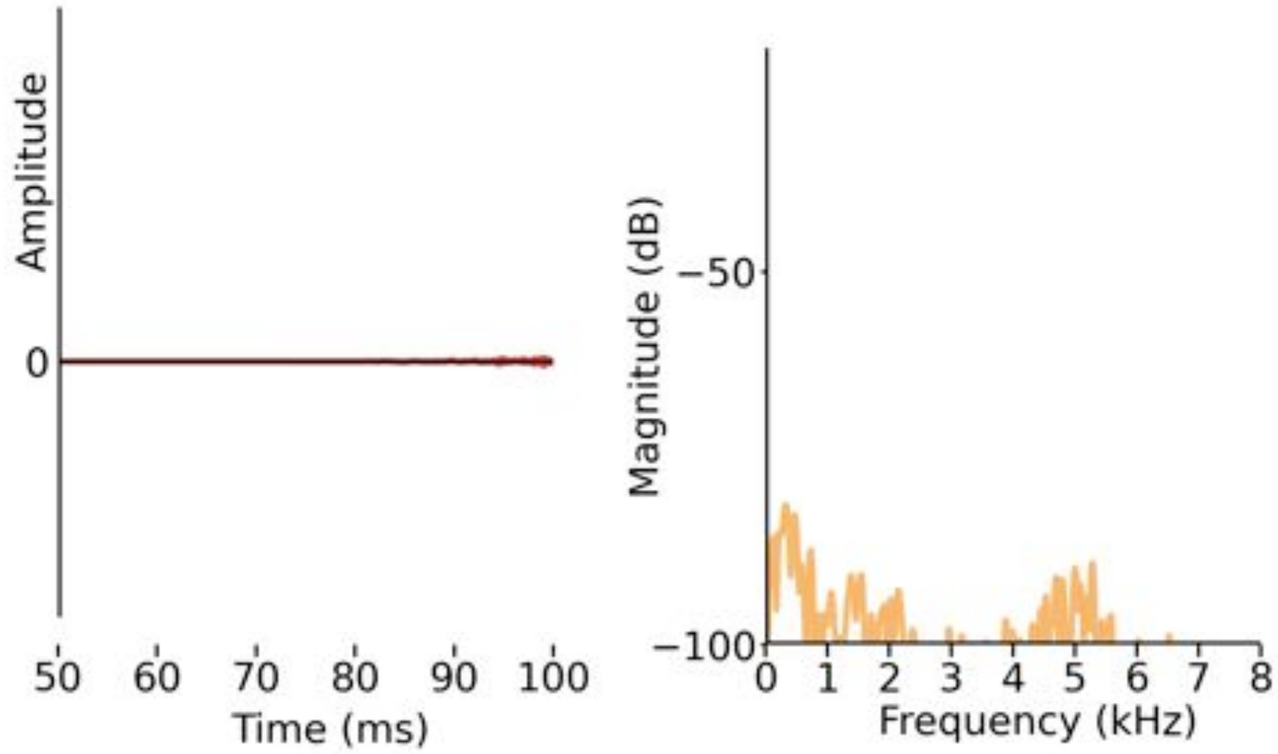


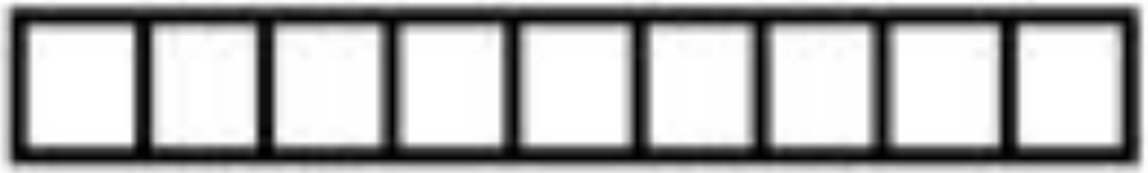
Recap: Filterbank features for automatic speech recognition

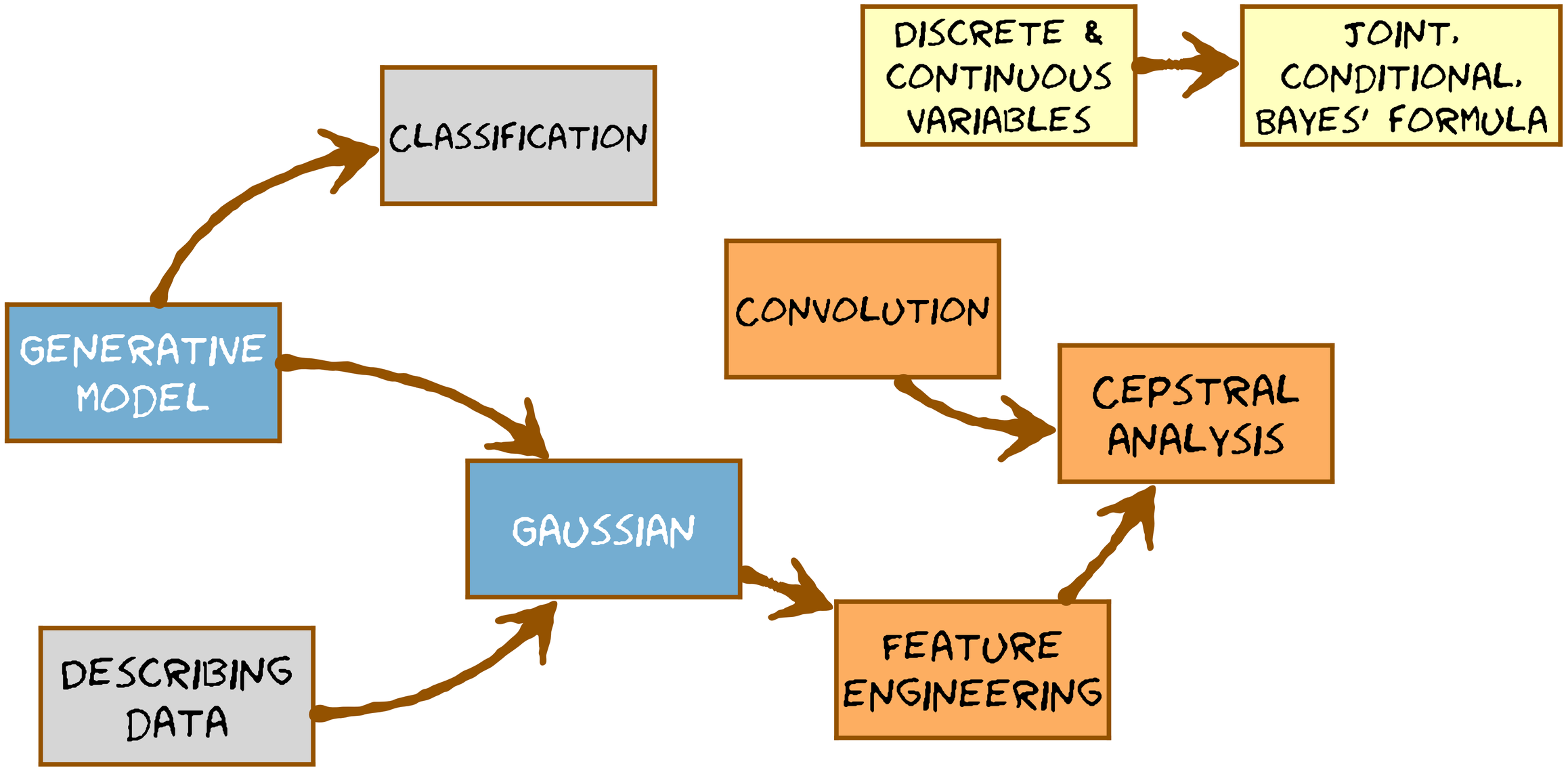




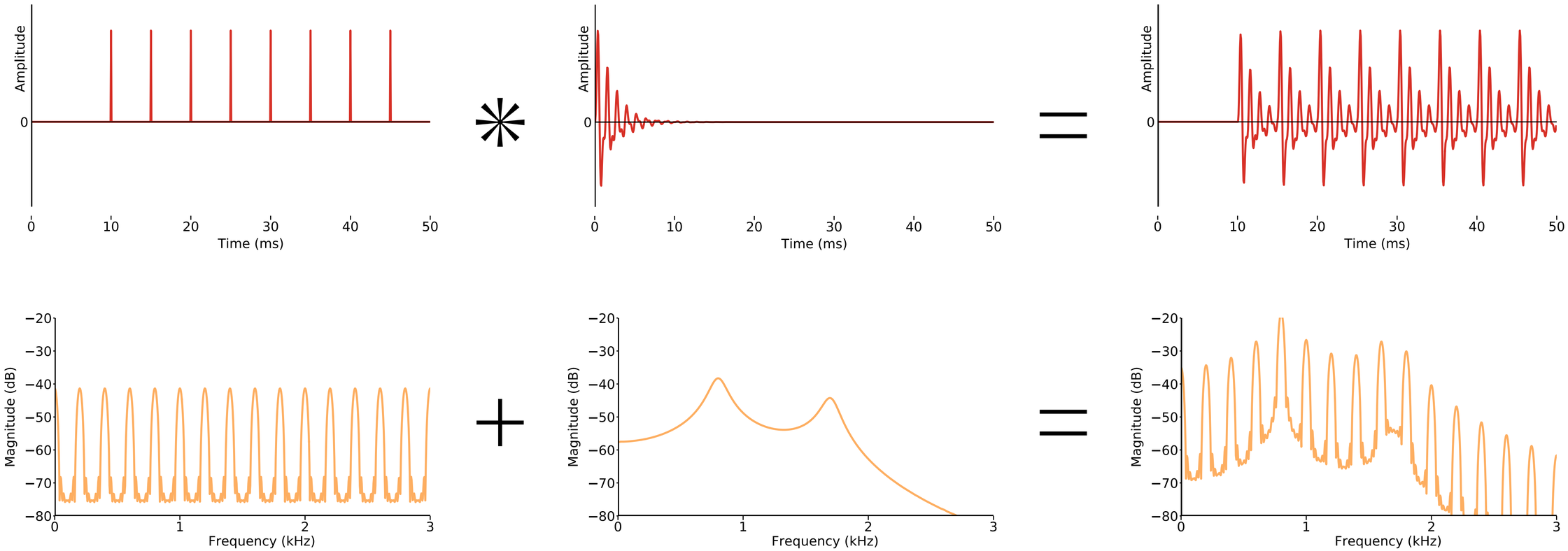
Correlation between features ⇨ need to model covariance



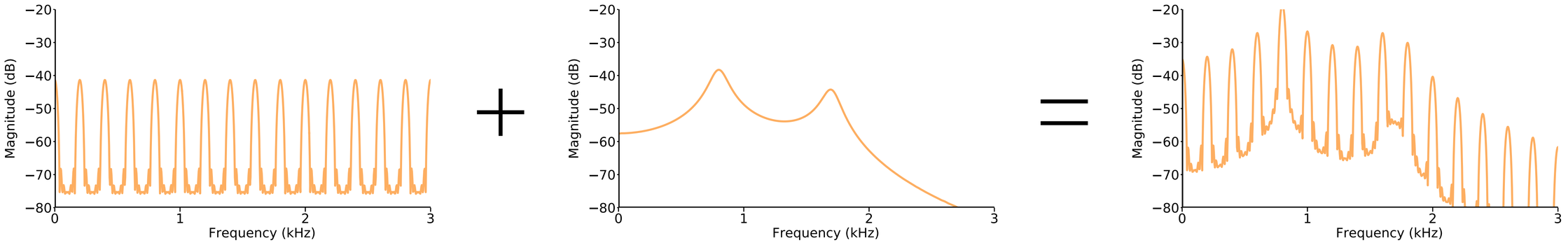



DISCRETE & CONTINUOUS VARIABLES
JOINT, CONDITIONAL, BAYES' FORMULA
CLASSIFICATION
GENERATIVE MODEL
CONVOLUTION
CEPSTRAL ANALYSIS
GAUSSIAN
FEATURE ENGINEERING
DESCRIBING DATA

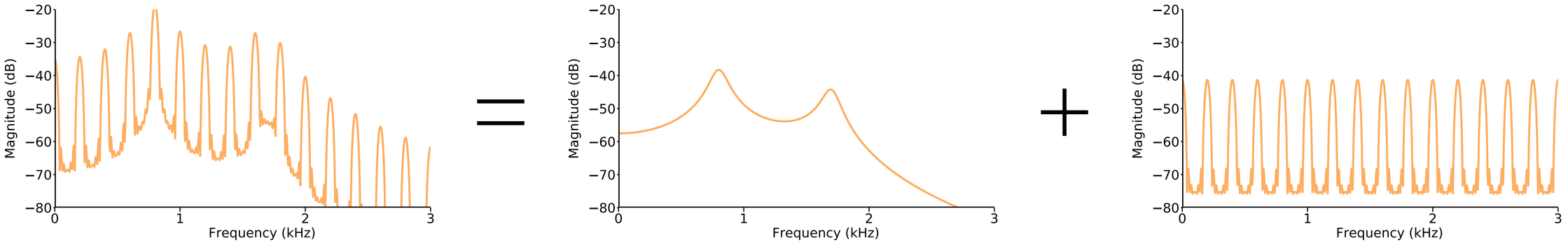
recap: convolution of waveforms = addition of log magnitude spectra



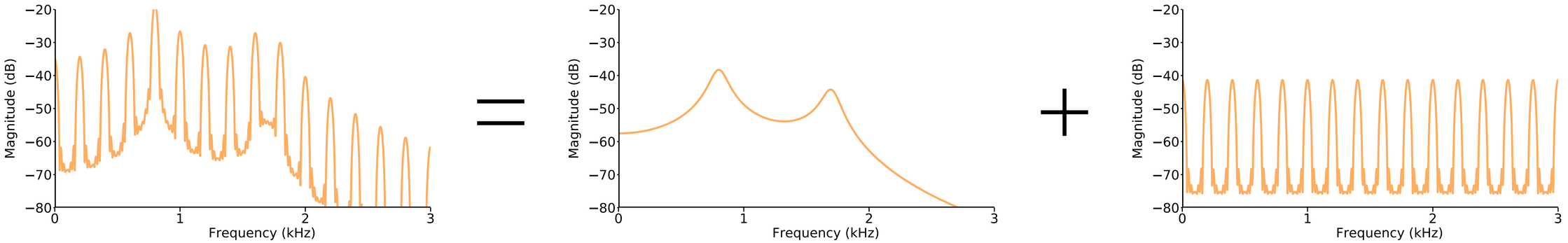
addition of log magnitude spectra



but we want to do this

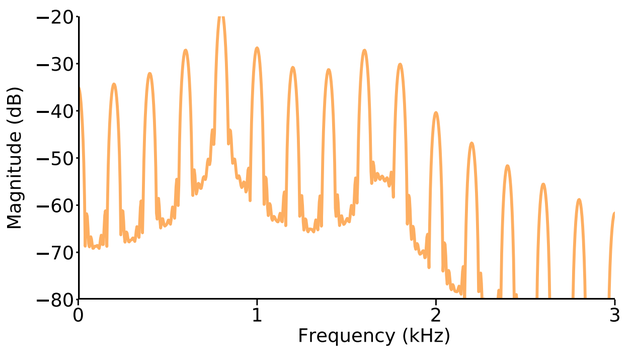


but we want to do this

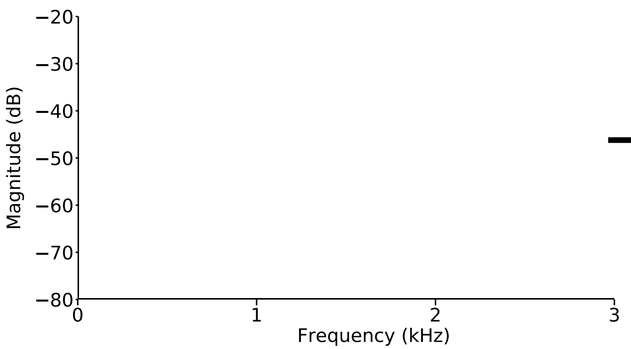


a more general expression that we can solve

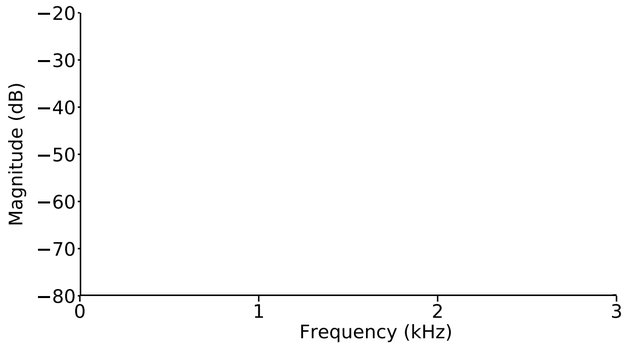
SERIES EXPANSION



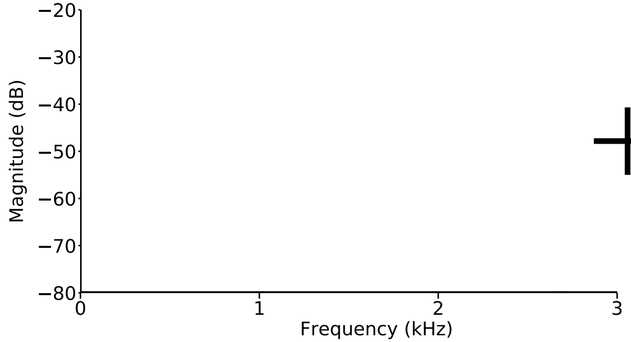
=



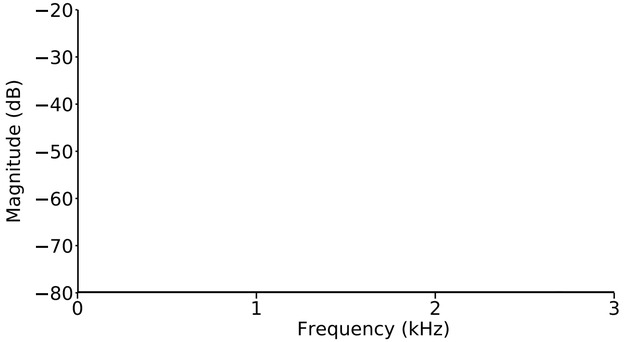
+



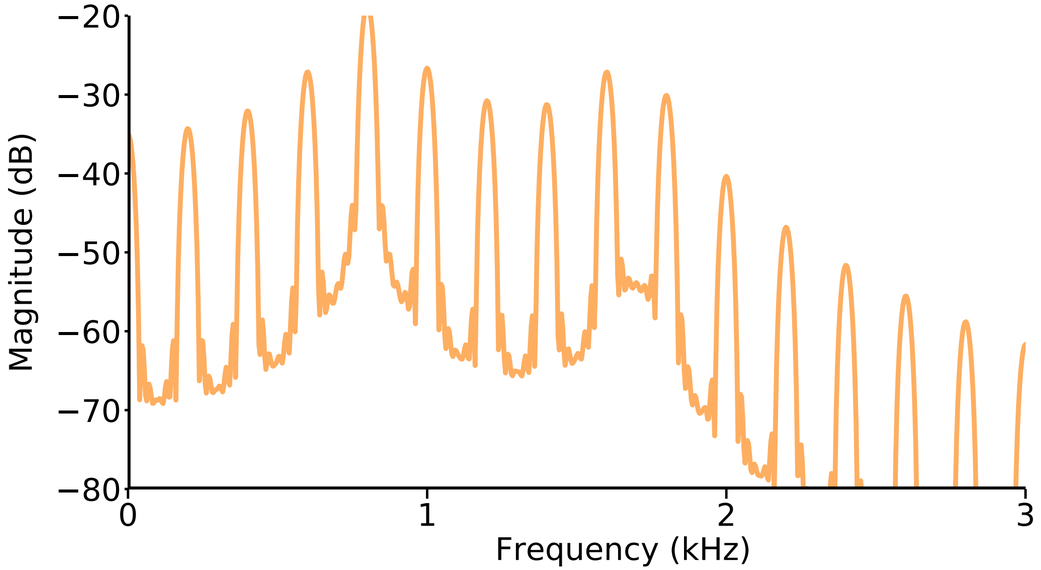
+



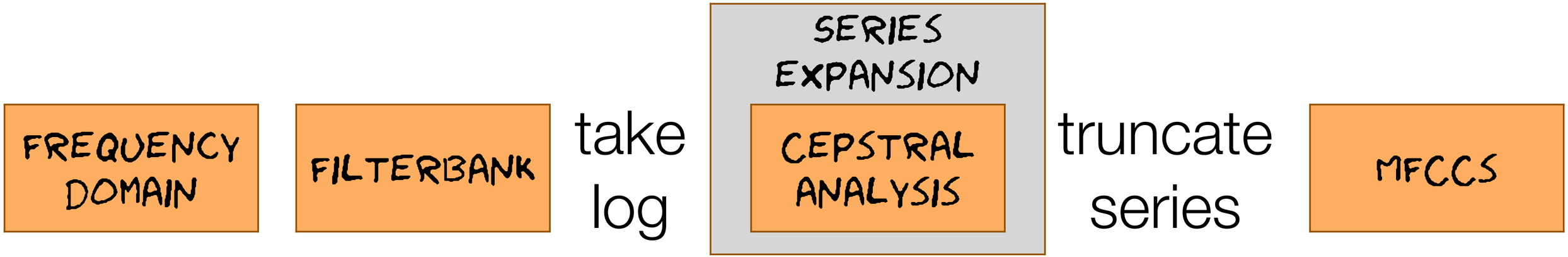
+



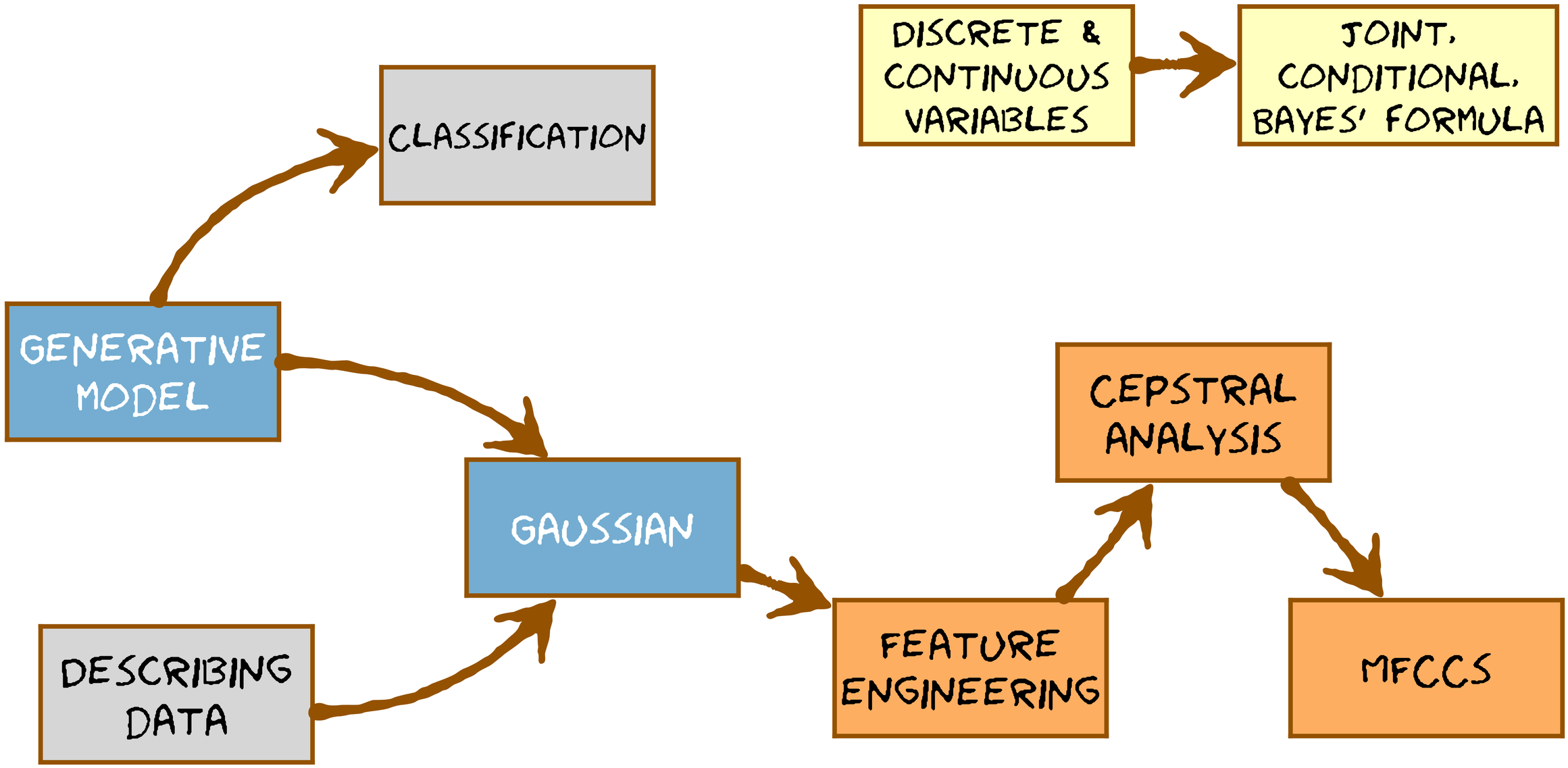
The spectrum and the cepstrum



Feature engineering: Mel Frequency Cepstral Coefficients



But did we remove covariance?


DISCRETE & CONTINUOUS VARIABLES
JOINT, CONDITIONAL, BAYES' FORMULA
CLASSIFICATION
GENERATIVE MODEL
GAUSSIAN
DESCRIBING DATA
CEPSTRAL ANALYSIS
FEATURE ENGINEERING
MFCCS

What next?

- From the Gaussian generative model to a model that **generates a sequence**
  - the Hidden Markov Model (HMM)

Module 9

- Deciding what to model
  - whole words ?
  - sub-word units ?

Module 10

- Connected speech

- Estimating the parameters of the HMM