Automating pattern discovery and the statistical process for regression

David Duvenaud[1], **James Robert Lloyd**[1], Roger Grosse[2],
Joshua B. Tenenbaum[2], Zoubin Ghahramani[1]

1: Department of Engineering, University of Cambridge, UK 2: Massachusetts Institute of Technology, USA

Massachusetts Institute of Technology

UNIVERSITY OF CAMBRIDGE

Computational and Biological Learning
University of Cambridge

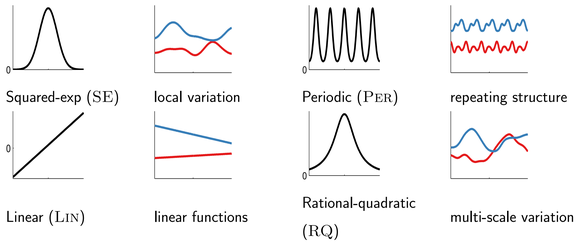
Data may often exhibit high level structure e.g. linearity, periodicity. . .

Display some 1d data sets - linear, quadratic, periodic, airline

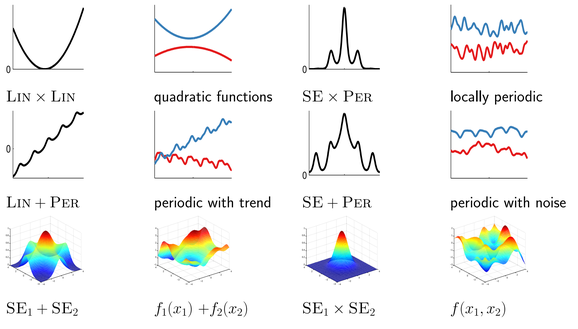
- Seemingly different regression methods would be required to model these data sets, and produce credible extrapolations
- . . .
- . . .

Gaussian process regression can model many structures with an appropriately chosen kernel

- The kernel encodes the inductive bias of the model i.e. the types of functions the model 'believes in'
- Below we depict standard base kernels, and examples of functions the model believes in (samples from the prior)



- Base kernels can be combined to create more complicated structural assumptions



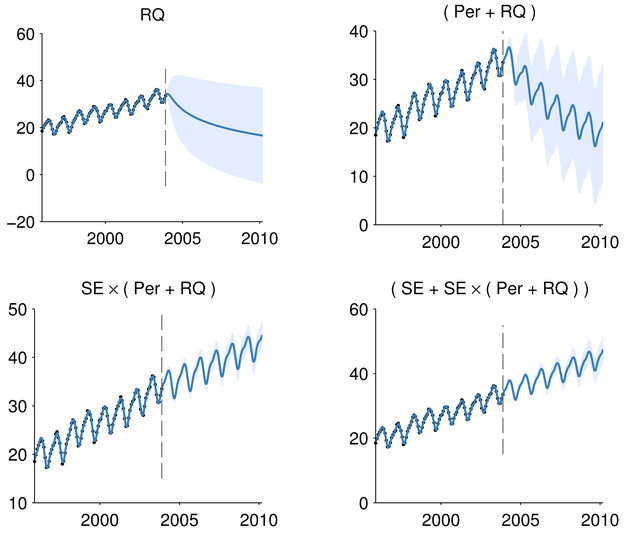
We consider all kernel expressions derived from a generative grammar. . .

- Constructing appropriate composite kernels has previously been the domain of Gaussian process experts
- We consider all algebraic expressions involving a small number of base kernels and the operations '$+$' and '$\times$'
- These operations are interpretable. . .
  - Addition of kernels corresponds to the addition of independent functions
  - Multiplication of kernels behaves like an 'AND' operation, combining properties of the base kernels
- . . . and produce a rich space including many standard statistical models

Bayesian linear regression	LIN
Bayesian polynomial regression	$\text{LIN} \times \text{LIN} \times \ldots$
Generalized Fourier decomposition	$\text{PER} + \text{PER} + \ldots$
Generalized additive models	$\sum_{d=1}^{D} \text{SE}_d$
Automatic relevance determination	$\prod_{d=1}^{D} \text{SE}_d$
Linear trend with deviations	$\text{LIN} + \text{SE}$
Linearly growing amplitude	$\text{LIN} \times \text{SE}$

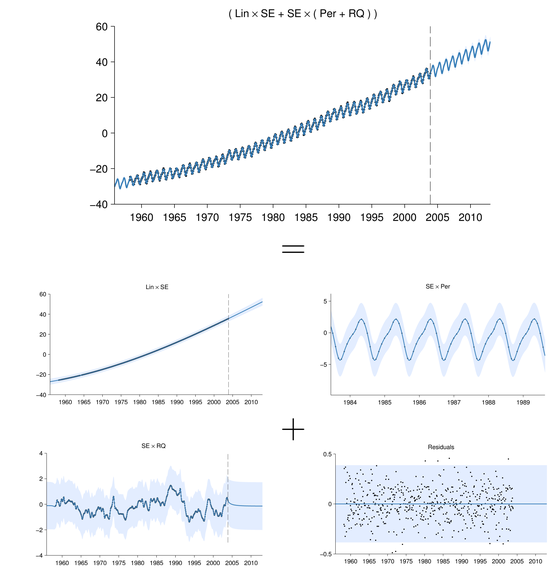
. . . which we search greedily, producing progressively better statistical models

Draw a picture of the search



Example: Mauna Loa CO_2 concentration

- By automatically inferring an appropriate kernel, we can also automatically decompose functions into additive components



Example: International airline passengers