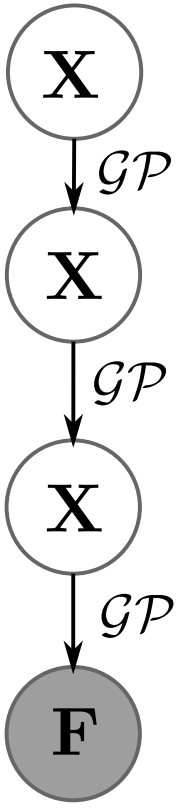
Deep Gaussian Processes

Jing Zhao
jzhao@cs.ecnu.edu.cn

Refer to the slide of Andreas Damianou at GPSS

Deep Gaussian processes

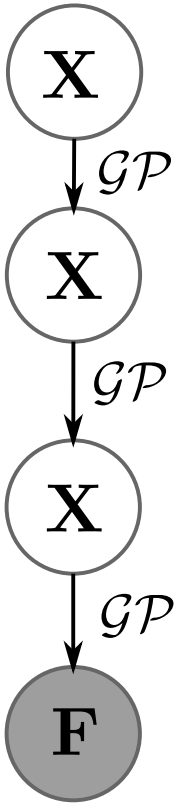


- Now recurse the stacked construction

$$f(\mathbf{x}) \rightarrow \text{GP}$$
$$f(x(\mathbf{t})) \rightarrow \text{stacked GP}$$
$$f(x_2(\mathbf{x}_1)) \rightarrow \text{stacked GP}$$
$$f(x(x(x \cdots (\mathbf{x}_1)))) \rightarrow \text{deep GP}$$

- The variational approximation changes only a little
- Uncertainty modelled "everywhere"!

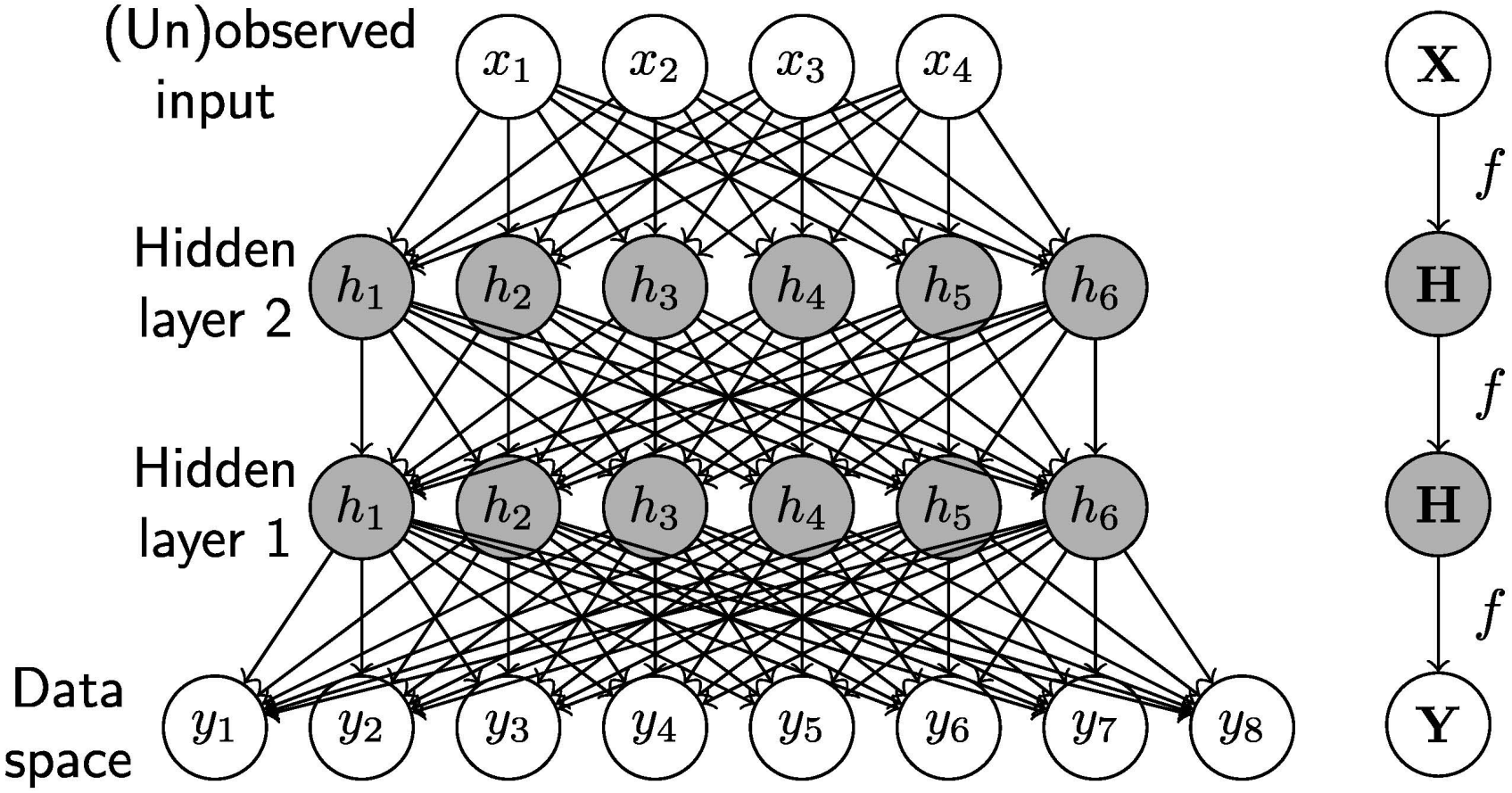
Deep Gaussian processes



- Now recurse the stacked construction

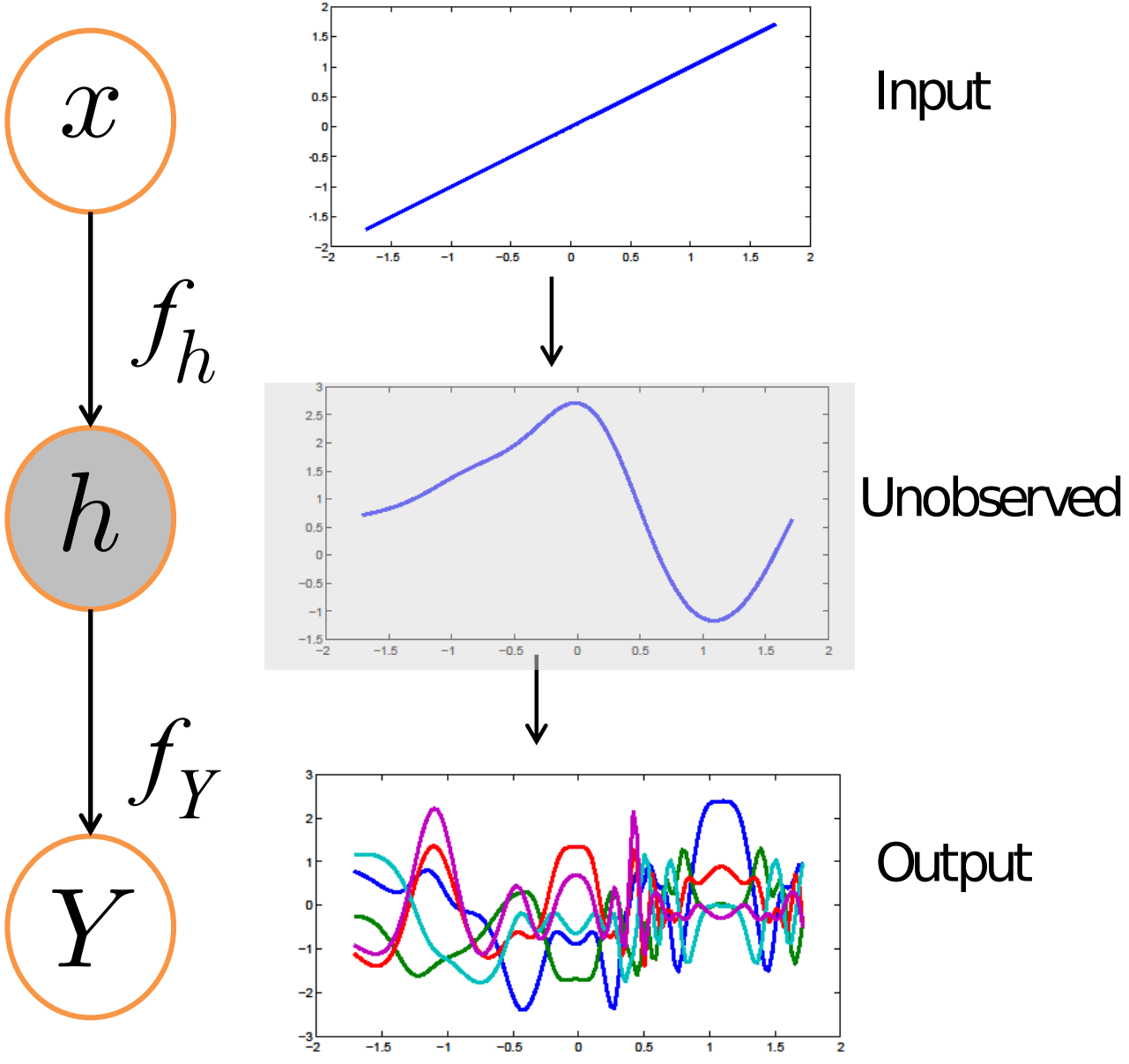
$$f(\mathbf{x}) \rightarrow \text{GP}$$
$$f(x(\mathbf{t})) \rightarrow \text{stacked GP}$$
$$f(x_2(\mathbf{x}_1)) \rightarrow \text{stacked GP}$$
$$f(x(x(x \cdots (\mathbf{x}_1)))) \rightarrow \text{deep GP}$$

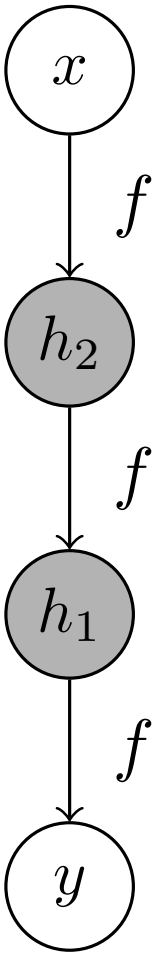
Deep learning



$$\mathbf{Y} = f(f(\cdots f(\mathbf{X})))$$

Sampling from a deep GP





- Joint $= p(y|h_1)p(h_1|h_2)p(h_2|x)$
- MAP optimization is extremely problematic because:
  - Dimensionality of hs has to be decided a priori
  - Prone to overfitting, if h are treated as parameters
  - Deep structures are not supported by the model's objective but have to be forced [Lawrence & Moore '07]

Regularization solution: approximate Bayesian framework

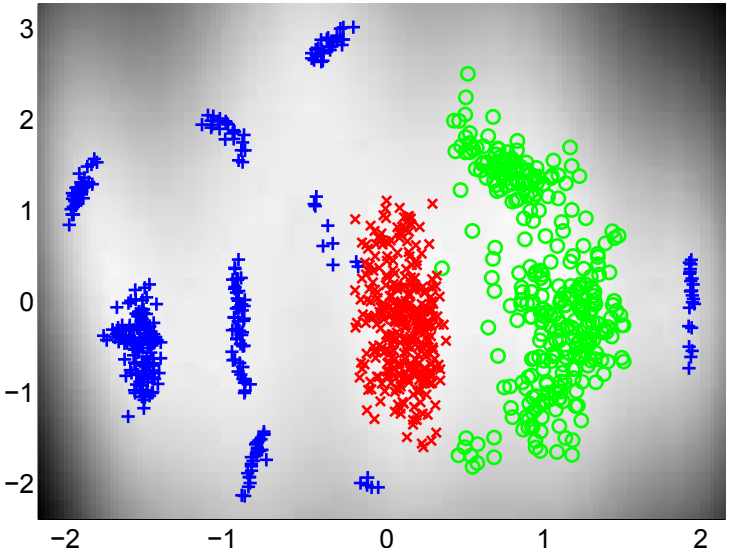
- Analytic variational bound $\mathcal{F} \leq p(y|x)$
  - Extend Titsias' method for *variational learning of inducing variables in Sparse GPs*.
  - *Approximately* marginalise out h
- Automatic structure discovery (nodes, connections, layers)
  - Use the Automatic / Manifold Relevance Determination trick
- ...

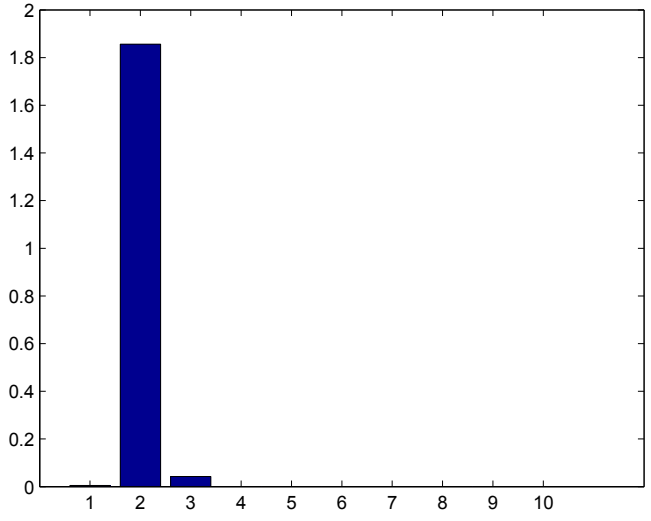
Automatic dimensionality detection

- Achieved by employing *automatic relevance determination (ARD)* priors for the mapping f.
- $f \sim \mathcal{GP}(\mathbf{0}, k_f)$ with:

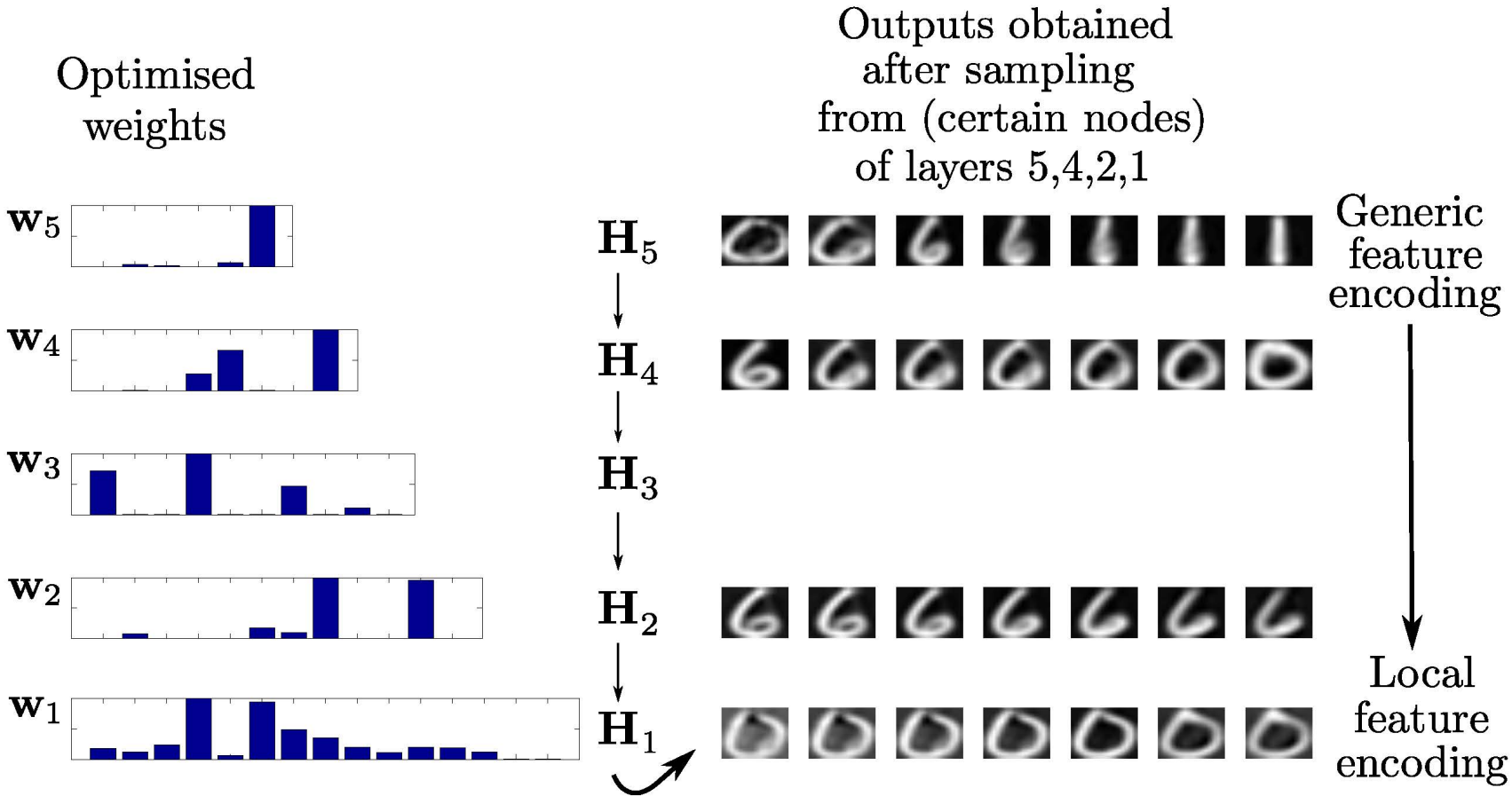
$$k_f\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sigma^2 \exp\left(-\frac{1}{2}\sum_{q=1}^{Q} w_q \left(x_{i,q} - x_{j,q}\right)^2\right)$$

- Example:





Deep GP: MNIST example

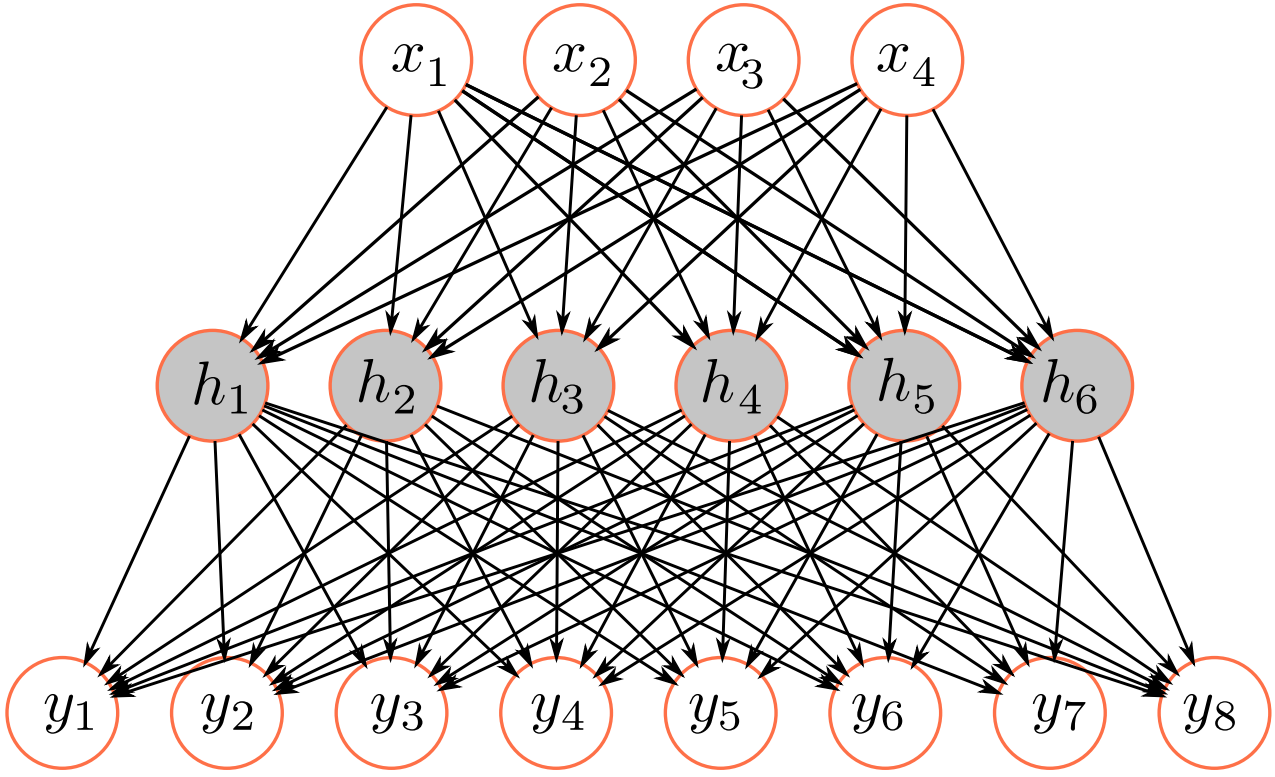


- http://staffwww.dcs.sheffield.ac.uk/people/A.Damianou/research/index.html#DeepGPs

Automatic structure discovery

Tools:

- ARD: Eliminate uncessary nodes/connections
- MRD: Conditional independencies
- Approximating evidence: Number of layers (?)



Automatic structure discovery

Tools:

- ARD: Eliminate uncessary nodes/connections
- MRD: Conditional independencies
- Approximating evidence: Number of layers (?)