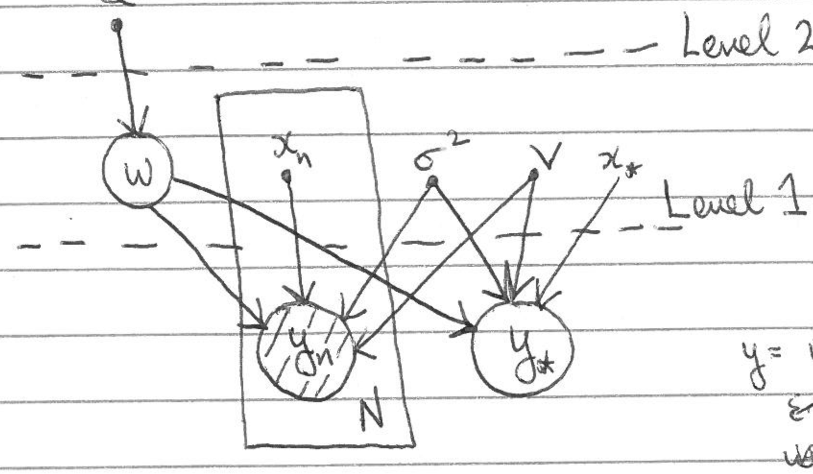
Bayesian Machine Learning Sep 1, 2016

Lecture 4



$y = \vec{w}^T \phi(x, v) + \varepsilon$

$\varepsilon \sim N(0, \sigma^2)$

$w \sim N(0, \alpha^2 I)$.

- From bottom (observed data, test data) to top (hyperparameter of our prior on w), to emphasise the generative process of data.
- This graphical representation helps us understand our models (joint distributions, conditional independence properties), and provides a natural mechanism for modification and construction.
- Captures a <u>causal</u> process by which the data are generated.

From the graph, the joint distribution is

$$p(y_*, y, w \mid x_*, X, \alpha^2, \sigma^2, \vec{v}) = \left[\prod_{n=1}^{N} p(y_n \mid x_n, w, \sigma^2, v)\right] p(w \mid \alpha)\, p(y_* \mid x_*, w, \sigma^2, v)$$

$$p(y_* \mid x_*, X, \vec{y}, \alpha^2, \sigma^2, v) \propto \int p(y_*, \vec{y}, \vec{w} \mid x_*, X, \alpha^2, \sigma^2)\, d\vec{w}$$