





#### **BAYESIAN REGRESSION: POSTERIOR DISTRIBUTION**

- Let's assume the regression weights have a Gaussian prior
   w ~ N(0, αl) and that the bias is zero
- The log posterior is

$$\log p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = -\frac{\beta}{2} \sum_{j=1}^{N} [t_j - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_j)]^2 - \alpha \mathbf{w}^T \mathbf{w} + const$$

• Hence the posterior is Gaussian

$$p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_{\mathsf{N}}, \mathbf{S}_{\mathsf{N}})$$

with mean and variance

$$\mathbf{m}_{\mathbf{N}} = \beta \mathbf{S}_{\mathbf{N}} \mathbf{\Phi}^T \mathbf{t}$$
$$\mathbf{S}_{\mathbf{N}}^{-1} = \alpha \mathbf{I} + \beta \mathbf{\Phi}^T \mathbf{\Phi}$$

### BAYESIAN REGRESSION: PREDICTIVE DISTRIBUTION

- Given the posterior, one can find the MAP estimate. However, in a fully Bayesian treatment, one makes predictions by integrating out the parameters via their posterior distribution.
- The predictive distribution

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{t}, \mathbf{w}, \alpha, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) d\mathbf{w}$$

is Gaussian

$$p(t|\mathbf{t}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m_N}^T \boldsymbol{\phi}(\mathbf{x}), \sigma_N^2(\mathbf{x}))$$

with mean  $\mathbf{m}_{\mathbf{N}}^{T} \boldsymbol{\phi}(\mathbf{x})$  and variance

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^T \mathbf{S}_{\mathbf{N}} \boldsymbol{\phi}(\mathbf{x})$$

# MARGINAL LIKELIHOOD

- We can find *α* and *β* by maximising the marginal likelihood:
   *p*(t|*α*, *β*)
- The log-marginal likelihood is:

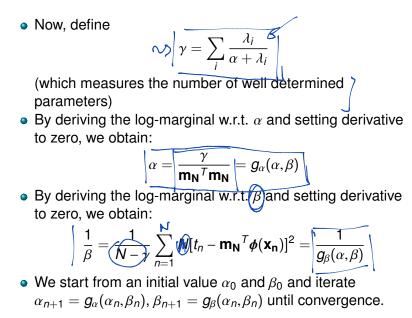
$$\sum_{\mathbf{n},\mathbf{n}} \log p(\mathbf{t}|\alpha,\beta) = \frac{M}{2} \log \alpha + \frac{N}{2} \log \beta - E(\mathbf{m}_{\mathbf{N}}) - \frac{1}{2} \log |\mathbf{S}_{\mathbf{N}}|^{-1} - \frac{N}{2} \log 2\pi$$
with
$$\sum_{\mathbf{n},\mathbf{n},\mathbf{n}} E(\mathbf{m}_{\mathbf{N}}) = \frac{\beta}{2} ||\mathbf{t} - \mathbf{\Phi}\mathbf{m}_{\mathbf{N}}||^{2} + \frac{\alpha}{2} \mathbf{m}_{\mathbf{N}}^{T} \mathbf{m}_{\mathbf{N}}$$

## **OPTIMISING THE MARGINAL LIKELIHOOD**

- We will present a fix-point algorithm: we will write the gradient equations equal to zero as fix-point equations and iterate until convergence.
- In taking the derivative w.r.t  $\alpha$  or  $\beta$ , the most challenging 44
- term is the log of the determinant of  $\mathbf{S_N}^{-1} = \alpha \mathbf{I} + \beta \mathbf{\Phi}^T \mathbf{\Phi}$  To deal with it, let  $\lambda_i$  be the eigenvalues of  $\beta \mathbf{\Phi}^T \mathbf{\Phi}$ , so that  $|\mathbf{S_N}^{-1}| = \prod_{i=0}^{M-1} (\alpha + \lambda_i)$ . TI elenvalue
- We then have that

• Moreover,  $\lambda_i$  are proportional to  $\beta$ , so that  $\partial \lambda_i / \partial \beta = \lambda_i / \beta$ 

# OPTIMISING THE MARGINAL LIKELIHOOD



# TASK 3

- Implement Bayesian regression, with type II likelihood optimisation of *α* and *β*.
- For the 1d non-linear dataset, use polynomial model of degree 12.
- Plot predictions and 95% confidence intervals, from the predictive distribution.
- For the 2d non-linear dataset, use the Gaussian functions models. How can we set the lengthscale γ?