Байесовский вывод:

- $p(D|\bar\theta)$ – likelihood

  – Gaussian noise:

$$\bar\varphi(\bar x)$$

$$p(y|\bar x,\bar w) = \mathcal{N}(y|\bar x^T\bar w, \sigma^2)$$

$$p(D|\bar w) = p(\bar y|X,\bar w) =$$

$$= \prod_n p(y_n|\bar x_n,\bar w) =$$

$$= \underline{\prod_n \mathcal{N}(y_n|\bar x_n^T\bar w, \sigma^2)}$$

- $\bar\theta_{ML} = \arg\max_\theta p(D|\bar\theta)$

$$\log p(D|\bar w) = \text{const} - \frac{1}{2\sigma^2}\sum_n (y_n - \bar x_n^T\bar w)^2 \longrightarrow \max$$

$$\Leftrightarrow \quad \sum_n (y_n - \bar x_n^T\bar w)^2 \xrightarrow[\bar w]{} \min$$

- $p(\bar\theta)$ – prior

$$p(\bar w) = \mathcal{N}(\bar w|\bar 0, \sigma_0^2 I)$$

$$\begin{pmatrix} \sigma_{01}^2 & & 0 \\ & \sigma_{02}^2 & \\ 0 & & \sigma_{0d}^2 \end{pmatrix}$$

$p(\bar\theta|D) \propto p(\bar\theta)p(D|\bar\theta)$ – posterior

$$p(\bar w) = \underline{\mathcal{N}(\bar w|\bar\mu_0, \Sigma_0)}$$

$$\log p(\bar\theta|D) = \text{const} + \log p(\theta) + \log p(D|\bar\theta)$$

$$\log p(\bar w|D) = \text{const} - \frac{1}{2\sigma^2}\sum_n (y_n - \bar x_n^T\bar w)^2$$

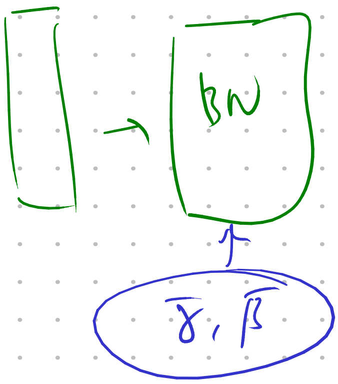
$$- \frac{1}{2}(\bar w - \bar\mu_0)^T\Sigma_0^{-1}(\bar w - \bar\mu_0)$$

$$p(\bar w|D) = \mathcal{N}(\bar w|\bar\mu_N, \Sigma_N)$$

- $p(y|\bar x, D) = \int \underline{p(y|\bar x,\bar w)\, p(\bar w|D)}\, d\bar w$ – predictive distribution

$$p(y|\bar x,\bar w)\,p(\bar w|D) = f(y,\bar x,D)\, \int \cancel{\mathcal{N}(\bar w|\mu', \Sigma')}\, d\bar w =$$

$$= \underline{p(y|\bar x, D)} = \mathcal{N}(y|\bar x^T\bar\mu_N, \sigma^2 + \ldots)$$



$$L_1 \xrightarrow{\theta} \min$$
$$L_2 \xrightarrow{\theta} \max$$

$$L_1 - \alpha L_2 \to \min$$

$$\frac{L_1}{L_2} \xrightarrow{\theta} \min$$

$$L = \lambda_1 L_1 + \lambda_2 L_2 + \dots + \lambda_{25} L_{25}$$