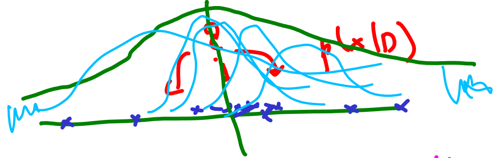


$$p(\theta|D) = \frac{p(\theta)\,p(D|\theta)}{p(D)}$$

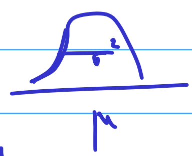
$\mu_{MAP} = \mu_N$

$$p(t|\bar{x}, D) = \int \underline{p(\theta)\,p(D|\theta)} \cdot p(t|\bar{x}, \theta)\, d\theta = \mathcal{N}(t \,|\, \mu_N^T \bar{x},\ \sigma^2 + \bar{x}^T \Sigma_N \bar{x})$$

$$p(\theta) \times p(D|\theta) \propto p(\theta|D)$$

$$\underline{p(\theta;\alpha)} \qquad\qquad \underline{p(\theta;\alpha')}$$

conjugate priors

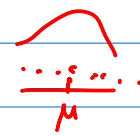


$\tau = 1/\sigma^2$

$$p(x|\mu,\tau) = \sqrt{\frac{\tau}{2\pi}} \cdot e^{-\frac{\tau}{2}(x-\mu)^2}$$

D

$$\boxed{p(\mu,\tau)} \times p(D|\mu,\tau) \propto p(\mu,\tau|D)$$



① $\tau = \text{const}$

$$p(\mu) \times p(D|\mu) \propto p(\mu|D)$$

$$\ln p(D|\mu) = \text{const} - \frac{\tau}{2}\sum_n (x_n - \mu)^2 \quad +$$

$$\ln p(\mu) = \frac{1}{2}\ln \tau_0 - \frac{1}{2}\ln 2\pi - \frac{\tau_0}{2}(\mu - \mu_0)^2$$

$$\ln \mathcal{N}(\mu|\mu_0, \tau_0) \quad -\frac{1}{2}\cdot(\tau_0 + \tau)\mu^2$$

$$\tau' = \tau_0 + \tau$$

② $\mu = \text{Const}$

$$p(\tau) \times p(D|\tau) \propto p(\tau|D)$$

$$\ln p(D|\tau) = \sum_n \ln p(x_n|\tau) =$$

$$= \sum_n \left[\frac{1}{2}\ln\tau - \frac{\tau}{2}(x_n - \mu)^2\right] + \text{const} =$$

$$= \frac{n}{2}\ln\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2 + \text{const}$$

$$\ln p(\tau) = (\alpha - 1)\ln\tau - \beta\cdot\tau + \text{const}$$

$$p(\tau|\alpha,\beta) = \frac{1}{Z}\tau^{\alpha-1}\cdot e^{-\beta\tau}$$

$$\ln p(\tau|D) = \text{const} + (\alpha_0 - 1)\ln\tau - \beta_0\tau + \frac{N}{2}\ln\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2$$

$$\alpha_N = \alpha_0 + \frac{N}{2} \qquad \beta_N = \beta_0 + \frac{1}{2}\sum_n (x_n - \mu)^2$$

③

$$p(\mu,\tau) \times p(D|\mu,\tau) \propto p(\mu,\tau|D)$$

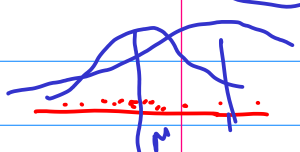
$$p(x|\mu,\tau) = -\frac{1}{2}\ln\tau - \frac{\tau}{2}(x-\mu)^2 + \text{const}$$

$\tau \cdot \mu$, $\tau \mu^2$

$$p(\mu,\tau) \stackrel{?}{=} \mathcal{N}(\mu|\mu_0,\tau_0)\cdot \text{Gam}(\tau|\alpha_0,\beta_0)$$

$$\ln \;=\; \ldots\mu^2\ldots\mu\ldots \quad -\ln\tau \ldots\tau$$

$$= c_1\mu^2 + c_2\mu + c_3\ln\tau + c_4\tau + c_5$$



$$p(\mu,\tau) \cancel{=} p(\mu)p(\tau)$$

$$p(\mu,\tau) = p(\tau)\,p(\mu|\tau) = \text{Gam}(\tau|\alpha_0,\beta_0)\cdot\mathcal{N}(\mu|\mu_0,\tau\cdot\lambda_0)$$

$$\ln p(\mu,\tau) = c + (\alpha_0 - 1)\ln\tau - \beta_0\tau + \frac{1}{2}\ln\tau - \frac{\tau\cdot\lambda_0}{2}(\mu-\mu_0)^2$$

$p(\theta) \times p(D|\theta) \propto p(\theta|D)$

$$p(\bar{x}|\bar{\eta}) = h(\bar{x}) \cdot g(\bar{\eta}) \cdot e^{\bar{\eta}^T \bar{u}(\bar{x})}$$

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \underbrace{-\frac{1}{2\sigma^2}}_{=} \cdot x^2 + \underbrace{\frac{\mu}{\sigma^2}}_{=} \cdot x \; \underset{=}{} \; -\frac{1}{2\sigma^2}\mu^2$$

$h(x) = 1$ $\qquad \bar{u}(x) = \begin{pmatrix} x^2 \\ x \end{pmatrix}$ $\qquad \bar{\eta} = \begin{pmatrix} -1/2\sigma^2 \\ \mu/\sigma^2 \end{pmatrix}$

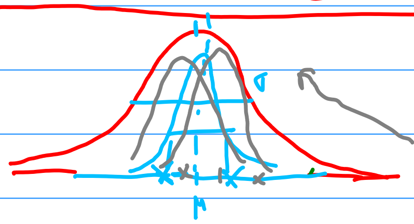
$g(\bar{\eta}) = \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot e^{-\frac{\mu^2}{2\sigma^2}}$

$\bar{u}(x) = \begin{pmatrix} -\frac{1}{2}x^2 \\ x \end{pmatrix}$ $\qquad \bar{\eta} = \begin{pmatrix} \tau \\ \mu\tau \end{pmatrix}$

$\{1, \text{p} \ \ 0, \text{opp}\}$

$$p(x|\theta) = \underline{\theta^x (1-\theta)^{1-x}} = e^{x \cdot \ln\theta + (1-x)\ln(1-\theta)} = (1-\theta) \cdot e^{x \cdot (\ln\theta - \ln(1-\theta))}$$

$$\boxed{\eta = \ln\frac{\theta}{1-\theta}}$$ log odds



$\mu_{ML} = \bar{x}$

$\sigma_{ML} = \frac{1}{\sqrt{N}} \cdot \sqrt{\Sigma(x-\bar{x})^2}$

$\sigma^* = \sqrt{\frac{1}{N-1}\Sigma(x-\bar{x})^2}$

$Gam(\tau|\ldots) \cdot \mathcal{N}(\mu|\ldots\tau)$

$\mathcal{N}(\bar{\mu}|\bar{\mu}_0, \ldots\Lambda\ldots) \cdot$

$\cdot \mathcal{W}(\Lambda|\ldots)$

wishart

$D = \{(\bar{x}_n, y_n)\}_{n=1}^N \qquad f: \bar{x} \mapsto y \qquad \bar{x}, y \sim p_{data}(\bar{x}, y) = \underline{p_d(\bar{x})} \cdot \underline{p_d(y|\bar{x})}$

exp. pred. error

$$\underline{EPE[f]} = \mathbb{E}_{\bar{x},y}(f(\bar{x}) - y)^2 \longrightarrow \min$$

$$\int (f(\bar{x}) - y)^2 p(\bar{x}, y)\, d\bar{x}\, dy =$$

$$f(\bar{x}) - y = f(\bar{x}) - \mathbb{E}_{y|\bar{x}} y + \mathbb{E}_{..} y - y$$

$$2\int (f - \mathbb{E}y) \cdot \left[\int (\mathbb{E}y - y) \underline{p(y|\bar{x})} dy\right] p(\bar{x}) d\bar{x} = 0$$

$$= \int (\underline{f - \mathbb{E}y + \mathbb{E}y - y})^2 p(\bar{x}, y) d\bar{x} dy = \int (f - \mathbb{E}y)^2 p(\bar{x}, y) d\bar{x} dy + \cancel{2\int (f - \mathbb{E}y)(\mathbb{E}y - y)\, \underline{p(\bar{x}, y)} d\bar{x} dy} +$$

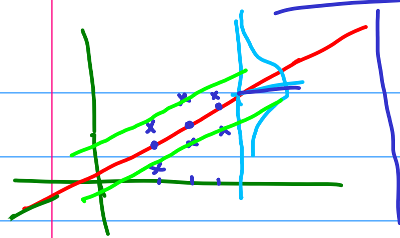
$$\text{"}p(\bar{x}) p(y|\bar{x})$$

$$(f - \mathbb{E}y)^2 + 2(f - \ldots)(\mathbb{E} - y) + (\mathbb{E}y - y)^2 \qquad + \int (\mathbb{E}y - y)^2 p(\bar{x}, y) d\bar{x} dy$$

$$= \boxed{\int (f - \mathbb{E}_{y|\bar{x}} y)^2 p(\bar{x}, y) d\bar{x} dy} + \boxed{\int (\mathbb{E}y - y)^2 p(\bar{x}, y) d\bar{x} dy}$$ noise

$$\boxed{\hat{f}(\bar{x}) = \mathbb{E}_{y|\bar{x}} y}$$ regression function

$$EPE[f] = \int (f(\bar{x};D) - \hat{f}(\bar{x}))^2 p(\bar{x},y)\, d\bar{x}\, dy + \int (E y - y)^2 p(\bar{x},y)\, d\bar{x}\, dy$$

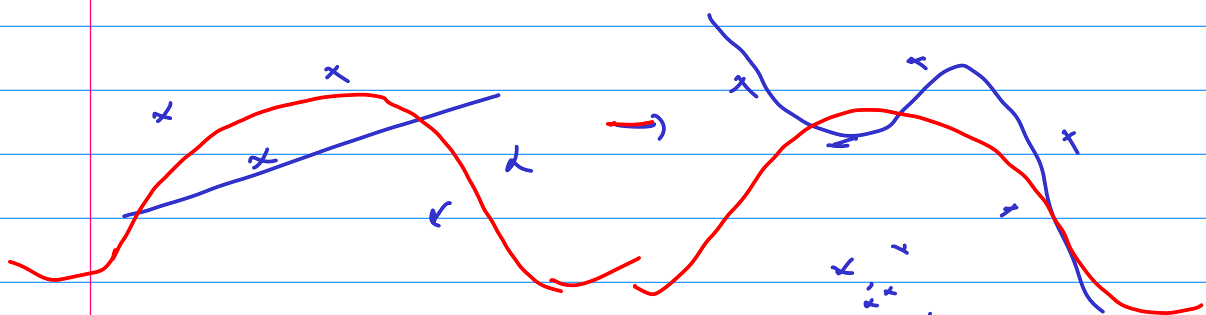


$$f_D - E_D f + E_D f - \hat{f}$$

$$\sim p(\bar{x},y)$$

$$\int (f_D - E_D f)^2 \ldots + 2 \int \cancel{(f - E f)(\ldots)}\, p(\bar{x},y)\, d\bar{x}\, dy + \int (E f - \hat{f})^2 \ldots$$

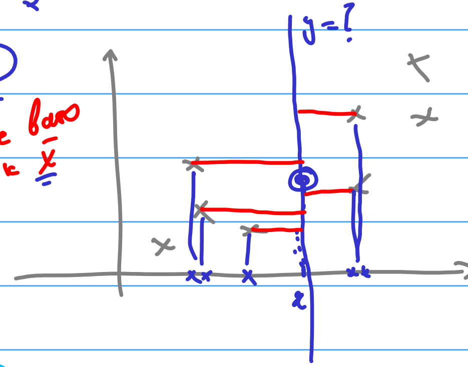
$$EPE[f] = \underbrace{\int (f_D(\bar{x}) - E_D f(\bar{x}))^2 p(\bar{x})\, d\bar{x}}_{\text{variance}} + \underbrace{\int (E_D f(\bar{x}) - \hat{f}(\bar{x}))^2 p(\bar{x})\, d\bar{x}}_{\text{bias}} + \underbrace{\int (\hat{f}(\bar{x}) - y)^2 p(\bar{x},y)\, d\bar{x}\, dy}_{\text{noise}}$$

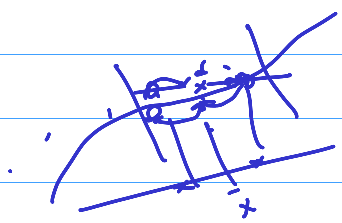


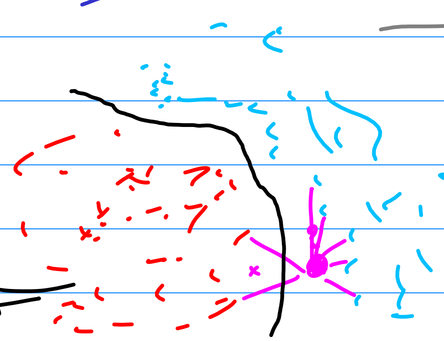
$$\hat{f}(\bar{x}) = E_{p(y|\bar{x})}\, y \approx \frac{1}{R}\sum_{r=1}^{R} y_r \approx \frac{1}{R}\sum y_r$$

$y_r \sim p(y|\bar{x})$

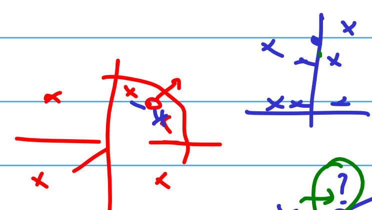
y^{x_r} close to $\bar{x}$

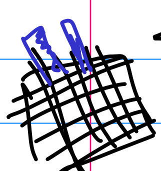


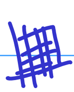


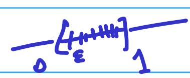


$$\int p(\theta)\, p(D|\theta)\, p(x|\theta)\, d\theta \approx$$









$d=1$ $\quad 1/\varepsilon$

$d=2$ $\quad 1/\varepsilon^2$

d $\quad 1/\varepsilon^d$

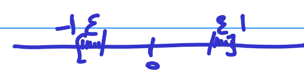
100–1000

$$\sum_{i=1}^{d} (x_i - y_i)^2$$

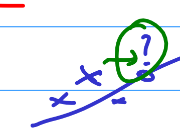
1/3

2/3

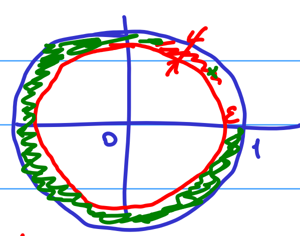
$d=1$



$$\frac{2\varepsilon}{2} = \varepsilon$$

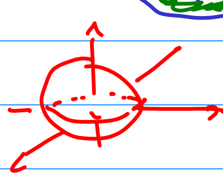


$d=2$



$$\frac{\pi - \pi(1-\varepsilon)^2}{\pi} = 2\varepsilon - \varepsilon^2$$

$d=3$



$$\frac{4/3\pi - 4/3\pi(1-\varepsilon)^3}{4/3\pi} = 3\varepsilon - 2\varepsilon^2 + \varepsilon^3$$

$$1 - (1-\varepsilon)^d \xrightarrow[d\to\infty]{} 1$$

d: