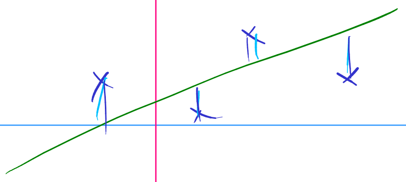


$$p(y|\bar{x},\bar{w}) = \mathcal{N}(y|\bar{w}^T\bar{x}, \sigma^2)$$

$$p(\bar{w}) = \mathcal{N}(\bar{w}|\underbrace{\bar{0}}_{\bar{\mu}_0}, \underbrace{\alpha \cdot I}_{\Sigma_0})$$

$$p(\bar{w}|D) \propto p(\bar{w})\,p(D|\bar{w}) \xrightarrow[\bar{w}]{} \max$$

$$\frac{1}{\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 + \alpha \cdot \bar{w}^T\bar{w} \xrightarrow[\bar{w}]{} \min$$

$$\ln p(\bar{w}|D) = \text{const} - \frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2 -$$

$$- \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma_0 - \frac{1}{2}(\bar{w}-\bar{\mu}_0)^T\Sigma_0^{-1}(\bar{w}-\bar{\mu}_0)$$

$$\shortparallel$$

$$\text{Const} - \frac{1}{2}(\bar{w}-\bar{\mu}')^T\Sigma'^{-1}(\bar{w}-\bar{\mu}')$$

$$\begin{cases} \Sigma'^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX \\ \boxed{\bar{\mu}' = \Sigma' \cdot \left(\Sigma_0^{-1}\bar{\mu}_0 + \frac{1}{\sigma^2}X^T\bar{y}\right)} \end{cases}$$

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

posterior

$$\mathcal{N}(a|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}\boxed{(a-\mu)^2}}$$

$$p(y|\bar{x},D) = \int p(y|\bar{x},\bar{w})\,p(\bar{w}|D)\,d\bar{w} =$$

$$\shortparallel$$

$$\int p(y,\bar{w}|\bar{x},D)\,d\bar{w}$$

$$= \int \mathcal{N}(y|\bar{w}^T\bar{x}, \sigma^2) \cdot \mathcal{N}(\bar{w}|\bar{\mu}', \Sigma')\,d\bar{w}$$

$$= \int e^{(\dots)}\,d\bar{w}$$

$$\ln(\dots) = -\frac{1}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(y - \bar{w}^T\bar{x})^2 - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma' -$$

$$- \frac{1}{2}(\bar{w}-\bar{\mu}')^T\Sigma'^{-1}(\bar{w}-\bar{\mu}')$$

$$\int \text{Const} \cdot \mathcal{N}(\bar{w}|\dots,\dots)\,d\bar{w} = \text{Const}(y)$$

$$e^{\dots y^2 \sim y}$$

$$p(y|\bar{x},D) = \mathcal{N}(y|\bar{w}^T\bar{x}, \sigma^2 + \bar{x}^T\Sigma'\bar{x})$$

$$p(y|D) = \int \underline{p(y|\bar\theta)}\, \underline{p(\bar\theta|D)}\, d\bar\theta = E_{p(\bar\theta|D)}[p(y|\bar\theta)]$$

$$\underline{\bar\theta^{(r)}} \sim p(\bar\theta|D)$$

$$r = 1, \ldots, R$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} p(y|\theta^{(r)})$$

$$E_{p(x)}[f(x)] \approx \frac{1}{R} \sum f(x^{(r)})$$

$$x^{(r)} \sim p(x)$$

$$\bar\mu' = \Sigma' \cdot \frac{1}{\sigma^2} X^T \bar y$$

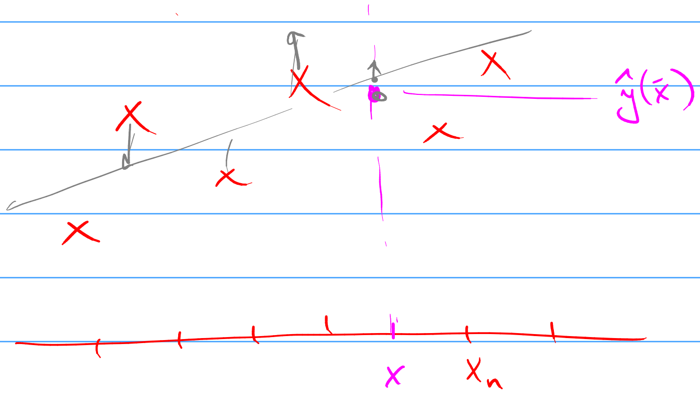
$$\begin{pmatrix} \vdots & & \\ x_n & & \\ \vdots & & \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \sum_{n=1}^{N} \underline{\bar x_n y_n}$$

$$\hat y(\bar x) = \bar\mu'^T \bar x = \bar x^T \bar\mu' = \bar x^T \cdot \Sigma' \cdot \frac{1}{\sigma^2} X^T \bar y = \bar x^T \cdot \Sigma' \cdot \frac{1}{\sigma^2} \cdot \sum_{n=1}^{N} \bar x_n y_n =$$

$$\underline{\hat y(\bar x) = \sum_{n=1}^{N} \left(\frac{1}{\sigma^2} \bar x^T \Sigma' \bar x_n \right) \cdot \underline{y_n}}$$

equivalent kernel

$$\hat y(\bar x) = \sum_{n=1}^{N} \underline{k(\bar x, \bar x_n)}\, y_n$$



$$p(\bar w) = N(\bar w | \bar 0, \alpha I_d)$$

$$1/\sigma_0^2$$

$$\sigma^2$$

$$\sigma_0^2$$

$$\boxed{\beta = \frac{1}{\sigma^2}}$$

precision

$$p(y|\bar x, \bar w) = N(y | \bar w^T \bar x, \beta)$$

$$\sqrt{\frac{\beta}{2\pi}} \cdot e^{-\frac{\beta}{2}(\bar w^T x - y)^2}$$

$$\ln p(\bar w|D) = \frac{N}{2} \ln \frac{\beta}{2\pi} - \frac{\beta}{2} \sum_n (y_n - \bar w^T \bar x_n)^2$$

$$+ \frac{d}{2} \ln \alpha - \frac{d}{2} \ln 2\pi - \frac{\alpha}{2} \bar w^T \bar w - \ln p(D)$$

$$p(\alpha, \beta | D) \rightarrow \max$$

$$\propto$$

$$\cancel{p(\alpha, \beta)} \cdot \underbrace{p(D|\alpha, \beta)}_{} \xrightarrow[\alpha, \beta]{} \max$$

$$\approx 1$$

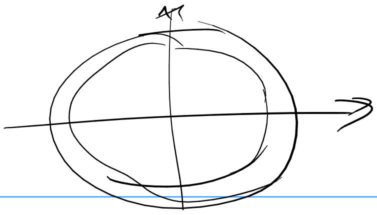
$$p(D|\alpha, \beta) = \int \underline{p(D|\bar w, \cancel{\alpha}, \beta)}\, \underline{p(\bar w|\alpha, \cancel{\beta})}\, d\bar w$$

$$= \int e^{\frac{N}{2} \ln \frac{\beta}{2\pi} - \frac{\beta}{2}(\_) + \frac{d}{2} \ln \alpha - \frac{d}{2} \ln 2\pi - \ldots}$$

$$\ln p(D|\alpha, \beta) \xrightarrow[\alpha, \beta]{} \max$$

Non-informative priors

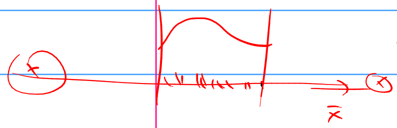
Curse of dimensionality



Statistical decision theory

$$L(y,\hat{y}) = (y-\hat{y})^2 \qquad (\bar{x},y) \in D$$

$$f: \bar{x} \mapsto \hat{y} = f(\bar{x})$$



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

$$EPE[f] = \mathbb{E}_{p(\bar{x},y)} L(y,\hat{y}) =$$

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

$$= \mathbb{E}_{p(\bar{x},y)} (y-f(\bar{x}))^2 =$$

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

$$= \int \left[\underbrace{\int (y-f(\bar{x}))^2 p(y|\bar{x})\, dy}_{} \right] p(\bar{x})\, d\bar{x}$$

$$f(\bar{x}) \to \min$$

$$\int (y-a)^2 q(y)\, dy$$

$$\mathbb{E}_{q(y)} \left[(y-a)^2 \right] \xrightarrow[a]{} \min$$

$$\hat{a} = \mathbb{E}_{q(y)}[y]$$

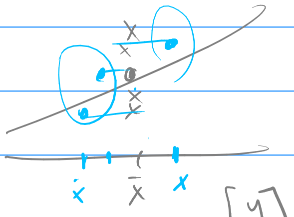
$$\boxed{\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y]}$$

regression function

$$\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y] \approx \frac{1}{R} \sum_{r=1}^{R} y^{(r)} \qquad y^{(r)} \sim p(y|\bar{x})$$

$$\approx \frac{1}{R} \sum y^{(r)}$$

$$y^{(r)} \in kNN(\bar{x})$$



$$\approx \mathbb{E}_{p(y|\bar{x})}[y]$$

$$L(y,\hat{y}) = \begin{cases} 1, & y \neq \hat{y} \\ 0, & y = \hat{y} \end{cases} = [y \neq \hat{y}]$$

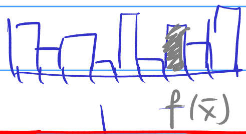
$$EPE[f] = \mathbb{E}_{p(\bar{x},y)} [L(y, f(\bar{x}))]$$

$$= \mathbb{E}_{p(\bar{x},y)} \left[[y \neq f(\bar{x})] \right] =$$

$$= \int \left(\underbrace{\sum_y [y \neq f(\bar{x})] p(y|\bar{x})}_{} \right) p(\bar{x})\, d\bar{x}$$

$$f(\bar{x}) \to \min$$

$$\boxed{\sum_y [y \neq f(\bar{x})] p(y|\bar{x})} \to \min$$



$$\hat{f}(\bar{x}) = \arg\max_y p(y|\bar{x})$$

$$L(y,\hat{y}) = \underbrace{\begin{pmatrix} 0 & 1 \\ 1000 & 0 \end{pmatrix} y}_{\hat{y}}$$

$$\sum_y L(y, f(x))\, p(y|\bar{x}) \to \min$$

$$-\hat{f}(\bar{x}) + \hat{f}(\bar{x}) \qquad ((y-\hat{f}) + (\hat{f}-f))^2$$

$$EPE[f] = \iint (y - f(\bar{x}))^2 p(\bar{x},y)\, d\bar{x}\, dy =$$

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

$$2\int\Big(\underbrace{\int(y-\hat{f}(\bar{x}))p(y|\bar{x})\,dy}\Big)(\hat{f}-f)p(\bar{x})\,d\bar{x}$$

$$\int(z-\mathbb{E}_p z)p(z)\,dz$$

Noise

$$=\overbrace{\int(y-\hat{f}(\bar{x}))^2 p(\bar{x},y)\,d\bar{x}\,dy}+\underbrace{\int(\hat{f}(\bar{x})\overset{-\mathbb{E}_D f+\mathbb{E}_D f}{-}f(\bar{x}))^2 p(\bar{x},y)\,d\bar{x}\,dy}$$

$$\mathbb{E}_D[f(\bar{x};D)] \qquad D\sim p(\bar{x},y)$$

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

$$\mathbb{E}PE[f]=\int(\hat{f}(\bar{x})-\mathbb{E}_D f(\bar{x}))^2 p(\bar{x},y)\,d\bar{x}\,dy \qquad \text{bias}^2$$

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

$$+\int(y-\hat{f}(\bar{x}))^2 p(\bar{x},y)\,d\bar{x}\,dy \qquad \text{noise}$$