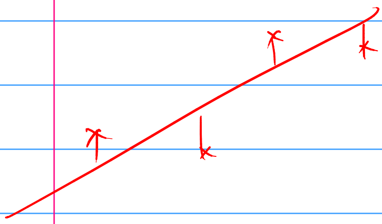
$$p(\theta|D) = \frac{p(\theta)\,p(D|\theta)}{p(D)} \to \max \quad \theta_{ML}$$

max θ_{MAP}

$$\underbrace{p(x|D)}_{\text{predictive distribution}} = \int p(x|\theta)\,\overbrace{p(\theta|D)}^{\text{posterior}}\,d\theta$$

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



$$p(\bar{y}|\bar{w}, X) = \prod_{n=1}^{N} p(y_n|\bar{w}, \bar{x}_n)$$

$$= \prod_n N(y_n|\bar{w}^T\bar{x}_n, \sigma^2)$$

$$p(\bar{w}) = N(\bar{w}|\bar{\mu}_0, \Sigma_0) =$$

$$= \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_0}} \cdot e^{-\frac{1}{2}(\bar{w}-\bar{\mu}_0)^T \Sigma_0^{-1}(\bar{w}-\bar{\mu}_0)}$$

$$(\bar{y}-X\bar{w})^T(\bar{y}-X\bar{w})$$

$$p(\bar{w}|\bar{y}, X) \propto p(\bar{w})\,p(\bar{y}|\bar{w}, X)$$

$$\ln p(\bar{w}|\bar{y}, X) = \text{const} - \frac{1}{2}(\bar{w}-\bar{\mu}_0)^T \Sigma_0^{-1}(\bar{w}-\bar{\mu}_0) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \bar{w}^T\bar{x}_n)^2$$

$$\Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^T X$$

$$\tilde{w}_{MAP} = \bar{\mu}_N = \dots$$

$$\mu_0, \Sigma_0 \rightsquigarrow \underset{(\bar{X}, \bar{y})}{D} \rightsquigarrow \bar{\mu}_N, \Sigma_N \longrightarrow D' \rightsquigarrow \bar{\mu}_{N'}, \Sigma_{N'}$$

$$p(y|\bar{x}, D) = \int p(y|\bar{x}, \bar{w})\,p(\bar{w}|D)\,d\bar{w} = E_{\boxed{p(\bar{w}|D)}}[p(y|\_)]$$

$$= \int N(y|\bar{w}^T\bar{x}, \sigma^2)\,N(\bar{w}|\bar{\mu}_N, \Sigma_N)\,d\bar{w} =$$

$$\dots e^{-\frac{1}{2\sigma^2}(y-\bar{w}^T\bar{x})^2} \quad \dots e^{-\frac{1}{2}(\bar{w}-\bar{\mu}_N)^T\Sigma_N^{-1}(\bar{w}-\bar{\mu}_N)}$$

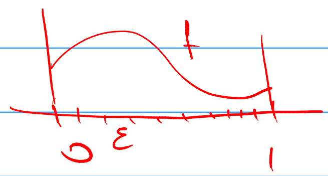
$$= C \cdot \underbrace{\int \mathcal{N}(\bar{w} \mid \dots)\, d\bar{w}}_{= 1}$$

$$\mathcal{N}(y \mid \dots, \dots)$$

$$p(y \mid \bar{x}, D) = \mathcal{N}\left(y \mid \bar{\mu}_N^T \bar{x},\ \sigma^2 + \bar{x}^T \Sigma_N \bar{x}\right)$$

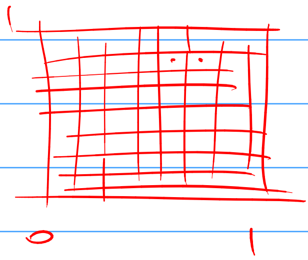
Curse of dimensionality

①



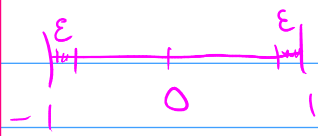
$$\frac{1}{\varepsilon} = N$$

$$\left(\frac{1}{\varepsilon^d}\right) = N^d$$



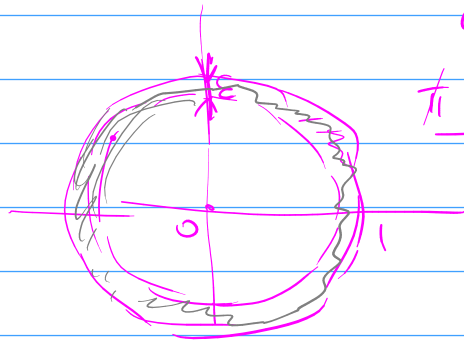
$$\frac{1}{\varepsilon^2}$$

②



$d = 1$

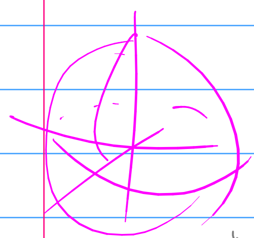
$$\varepsilon = 1 - (1 - \varepsilon)$$



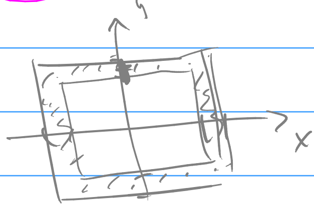
$d = 2$

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

$d = 3$

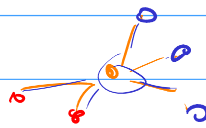


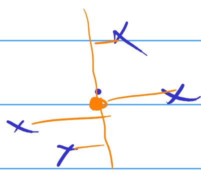
$$\frac{\frac{4}{3}\pi - \frac{4}{3}\pi(1-\varepsilon)^3}{\frac{4}{3}\pi} = 1 - (1-\varepsilon)^3$$

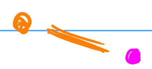


ε

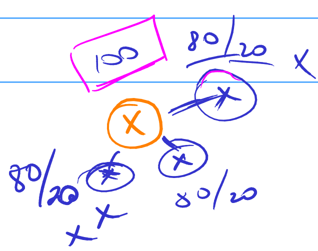
$$1 - (1 - \varepsilon)^d \xrightarrow[d \to \infty]{} 1$$

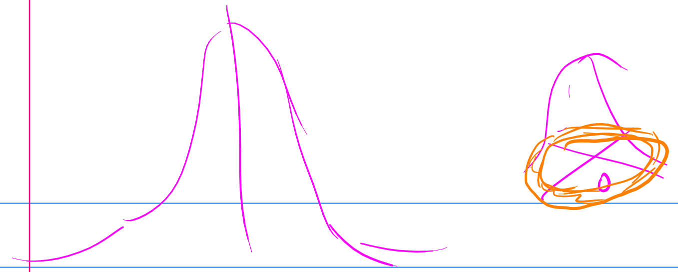




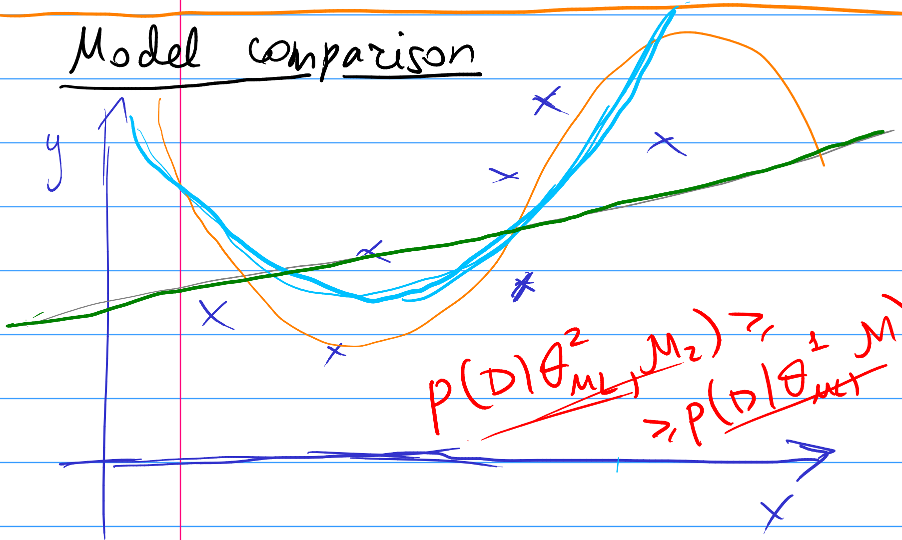


$$\sqrt{\sum (x_i - x_i')^2}$$





Model comparison



$M_1 \quad \bar{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$

$M_2 \quad \bar{w} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$

M_3

$p(D|\theta^2_{ML}, M_2) \geq p(D|\theta^1_{ML}, M_1)$

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

$$p(M_i|D) \propto p(M_i)\,\boxed{p(D|M_i)} \to \max$$

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

$p(D|\theta^*, M_i)\,\Delta\theta_{post}$

$0,$

$$p(D|M_i) = \int p(\theta|M_i)\,p(D|\theta, M_i)\,d\theta$$

$$p(D|M_i) = \int \frac{1}{\Delta\theta_{pri}} \cdot p(D|\theta, M_i)\,d\theta$$

$\theta^* = \theta_{ML} = \theta_{MAP}$

$$= \frac{p(D|\theta^*, M_i)}{\Delta\theta_{pri}} \int 1\,d\theta = p(D|\theta^*) \cdot \frac{\Delta\theta_{post}}{\Delta\theta_{pri}}$$

$\Delta\theta_{post}$

$$\boxed{BIC(M_i) = \ln p(D|\theta^*_1, M_i) - \frac{1}{2} M \ln N}$$

$$\ln p(D|M_i) \approx \ln p(D|\theta^*, M_i) - M \cdot \ln \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

$p(D|M)$

$\bar{x} \in \mathbb{R}^p \xrightarrow{f} y \in \mathbb{R} \qquad p(\bar{x}, y) \qquad p(y|x) = \frac{p(\bar{x}, y)}{p(\bar{x})}$

$L(y, f(\bar{x})) = (y - f(x))^2$

$$\underbrace{EPE[f]}_{\to \min} = \mathbb{E}_{\bar{x}, y}[L(y, f(\bar{x}))] = \iint (y - f(\bar{x}))^2 \underbrace{p(\bar{x}, y)}_{= p(\bar{x}) p(y|\bar{x})} d\bar{x}\, dy$$

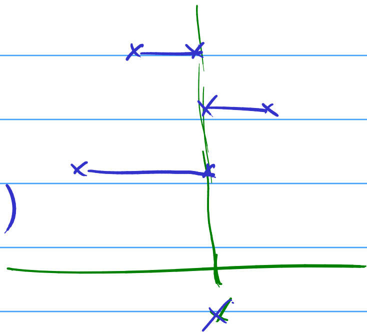
$$EPE[f] = \int \Big[\underbrace{\int (y - f(\bar{x}))^2 p(y|\bar{x})\, dy}_{} \Big] p(\bar{x})\, d\bar{x}$$

$$\int (x-a)^2 p(x)\, dx \underset{a}{\to} \min \qquad a = \mathbb{E}_p[x]$$

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

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

k-NN

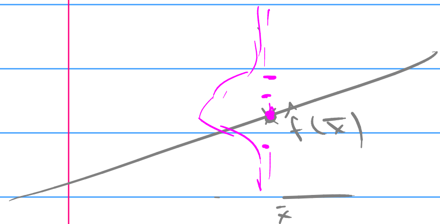


$$\underline{EPE[f]} = \iint (y - f(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy =$$

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

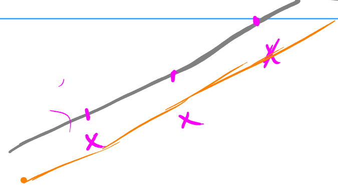
$$= \iint (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy + \cancel{2 \iint (y - \hat{f})(\hat{f} - f) p(\bar{x}, y)\, d\bar{x}\, dy} + \iint (\hat{f} - f)^2 p(\bar{x}, y)\, d\bar{x}\, dy$$

$$2 \int \Big[\underbrace{\int (y - \hat{f}(\bar{x})) p(y|\bar{x})\, dy}_{= 0} \Big] \ldots$$



$$= \underbrace{\mathbb{E}[(y - \hat{f}(\bar{x}))^2]}_{\text{NOISE}} + \mathbb{E}[(\hat{f}(\bar{x}) - f(\bar{x}))^2]$$

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



$$E_D f(\bar{x}; D)$$

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

Bias-Variance-Noise decomposition

$$EPE[f] = \underbrace{\iint (y - \hat{f}(\bar{x}))^2\, d\bar{x}\, dy}_{} \quad \text{NOISE}$$

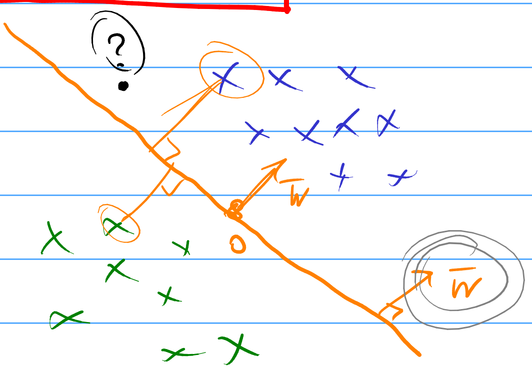
$$+ \underbrace{\iint (f(\bar{x}) - E_D f(\bar{x}; D))^2\, d\bar{x}\, dy}_{} \quad \text{VARIANCE}$$

$$+ \underbrace{\iint (E_D f(\bar{x}; D) - \hat{f}(\bar{x}))^2\, dx\, dy}_{} \quad \text{BIAS}$$

Classification

$y = \bar{w}^T \bar{x}$ $\qquad$ $y(\bar{x}) = 0$

$y(\bar{x}) = \bar{w}^T \bar{x}$



$y_2 = y_3 > y_1$

$y_1 = y_2 > y_3$

$C(\bar{x}) = \arg\max \{ y_1(\bar{x}), y_2(\bar{x}), y_3(\bar{x}) \}$