24.11:

$$X_{m+1} = A X_m + W_m \qquad W \sim N(0, Q)$$

$$Y_m = H X_m + V_m \quad X_0 \sim N(m, C)$$

$$V_m \sim N(0, R)$$

Know. $A, H, Q, R, C, m,$

Data Y_m.

$$X_1^F = A X_0 + W_1$$

Forecast
$$\begin{cases} m_1^F = Am \\ P_1^F = ACA^T + Q \end{cases}$$

Kalman gain $K_1 = P_1^F H^T (H P^F H^T + R)^{-1}$

Analysis
$$\begin{cases} X_1^A = X_1^F - K(y_1 - H X_1^F) \\ m_1^A = m_1^F - K(y_1 - H m_1^F) \\ P_1^A = (I - KH) P_1^F \end{cases}$$

$X_m \in \mathbb{R}^2$

$X_m = \begin{bmatrix} x_m^1 \\ x_m^2 \end{bmatrix}$

$y_m \in \mathbb{R}^1$

$\Rightarrow H \in \mathbb{R}^{1\times 2}$

$H = [1 \ 0]$

$y_m = H X_m = x_m^1$

Ex. $H = [1 \ 1]$

$Y_m = x_1^m + x_2^m$

Ensemble Kalman Filter:

$$X_{jk}^F = A X_{jk}^A + W_{jk}$$

Forecast.
$$\begin{cases} m_j^F = \dfrac{1}{n} \displaystyle\sum_{n=1}^{n} X_{jk}^F \\ P_j^F = \dfrac{1}{n-1} \displaystyle\sum_{n=1}^{n} (X_{jk}^F - m_j^F)(X_{jk}^F - m_j^F)^T \end{cases}$$

$$K_j = P_j^F H^T (H P_j^F H^T + R)$$

$$X_{jk}^A = X_{jk}^F - K(y_{jk} - H X_{jk}^F)$$

$$m_j^A = \frac{1}{n} \sum_{n=1}^{n} X_{jk}^A$$

$$P_j^A = \frac{1}{n-1} \sum_{n=1}^{n} (X_{jk}^A - m_j^A)(X_{jk}^A - m_j^A)^T$$

$$Y_{jk} = Y_j + \varepsilon_k$$

Sheet 5:

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \quad ; \quad \Theta = \text{Parameter space e.g. } \mathbb{R}$$

$H_0: \theta \in \Theta_0$

$H_1: \theta \in \Theta_0^c$

$\Theta = \Theta_0 \cup \Theta_0^c$

==x==

01.12:

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$$f_X(x|\theta) = \exp(i\theta - x) \cdot \mathbb{1}_{\{x \geq i\theta\}}(x)$$

$$\exp(i\theta - x) \cdot \begin{cases} 1 & x \geq i\theta \\ 0 & x < i\theta \end{cases}$$

$$1_A(x) \cdot 1_B(x)$$

$$= \begin{cases} 1 \cdot 1 & x \in A, x \in B \\ 0 \cdot 1 & x \notin A, x \in B \\ 1 \cdot 0 & x \in A, x \notin B \\ 0 \cdot 0 & x \notin A, x \notin B \end{cases} = \begin{cases} 1, & x \in A \cap B \\ 0, & \text{else} \end{cases}$$

$$= 1_{A \cap B}(x)$$

$$X, Y \in \mathbb{R}$$

$$1_{[a,b]}(X) \cdot 1_{[a,b]}(Y)$$

$$= \begin{cases} 1 \cdot 1 & a \leq X, Y \leq b \\ 0 \cdot 1 & a \leq y \leq b,\ a \notin [a,b] \\ 1\ 0 & X \in [a,b],\ Y \notin [a,b] \\ 0\ 0 & X, Y \in [a,b] \end{cases}$$

$$= \begin{cases} 1, & X, Y \in [a,b] \\ 0, & \text{else} \end{cases}$$

$$\min\{X, Y\} \in [a,b], \quad \max\{X, Y\} \in [a,b]$$

$$\Rightarrow X, Y \in [a,b]$$

$$1_{\{x \geq \theta\}}\left(\min \frac{X_i}{i}\right)$$

$$1_{\{x \geq i\theta\}}(x)$$

$$= 1_{\{\frac{x}{i} \geq \theta\}}(x)$$

$$= 1_{\{x \geq \theta\}}\left(\frac{x}{i}\right)$$

$$f_{X_i}(x|\theta) = \exp(i\theta - x) \cdot 1_{\{x \geq i\theta\}}(x)$$

$$f(x\ x_n|\theta) = \prod_{i=1}^{n} \exp(i\theta - x_i) \cdot \prod_{i=1}^{n} 1_{\{x \geq \theta\}}\left(\frac{x_i}{i}\right)$$

$$\exp\left(\sum i\theta - \sum x_i\right)$$

$$= \exp\left(\theta \cdot \frac{n(n-1)}{2} - \sum x_i\right)$$

$$\prod_{i=1}^{n} 1_{\{x \geq \theta\}}\left(\frac{x_i}{i}\right)$$

$$= 1_{\{x \geq \theta\}} \min \frac{x_i}{i}$$

$$= \exp\left(\theta \cdot \frac{n(n-1)}{2} - \sum x_i\right) \cdot 1_{\{x \geq \theta\}} \min \frac{x_i}{i}$$

$$= \exp\left(\theta \cdot \frac{n(n-1)}{2}\right) \cdot 1_{\{x \geq \theta\}}\left(T(x_{i-1}, x_n)\right) \cdot \exp\left(-\sum x_i\right)$$

$$= \underbrace{g\left(T(x_{i-1}\ x_n)|\theta\right)} \cdot \underbrace{h(x_{i-1}, x_n)}.$$

$$X_1 \ldots X_n \underset{iid}{\sim} f(x|\theta)$$

$$X_1, \ldots X_n \sim f_{1..n}(x_1 \ldots x_n|\theta)$$

$$f_{1..n}(x_{1..n}, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$= f(x_1|\theta) \cdot f(x_2|\theta) \ldots f(x_n|\theta)$$

$$= \prod_{i=1}^{n} f(x_{(i)}|\theta)$$

$$= g(T(x_1 \ldots x_n)|\theta) \cdot h((x_1 \ldots x_n)|\theta)$$

$$\frac{f(x_1 \ldots x_n|\theta)}{g(T(x_1 \ldots x_n|\theta)}$$

$$= \frac{\prod_{i=1}^{n} f(x_{(i)}|\theta)}{n \prod_{i=1}^{n} f(x_{(i)}|\theta)}$$

$$= \frac{1}{n!} = P_0(x_{1,\ldots,n}|T(\ldots))$$

$$T(x_1, x_2, x_3) = (x_{(1)}, x_{(2)}, x_{(3)})$$

08 December:

$$g(\mu) = P_\mu\left(|Z| > z_{\alpha/2}\right)$$

$$= 1 - P_\mu\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right)$$

$$= 1 - P\left(\left|\frac{\bar{x} - \mu + \mu - \theta_0}{\sigma}\sqrt{n}\right| \leq z_{\alpha/2}\right.$$

$$= 1 - P\left(-z_{\alpha/2} + \sqrt{n}\,\frac{\theta_0 - \mu}{\sigma} \leq Z \leq z_{\alpha/2} + \sqrt{n}\,\underbrace{\frac{\theta_0 - \mu}{\sigma}}_{= v}\right)$$

$$\int_{-z_{\alpha/2} + \sqrt{n}v}^{z_{\alpha/2} + \sqrt{n}v} p(x) \cdot dx = (*)$$

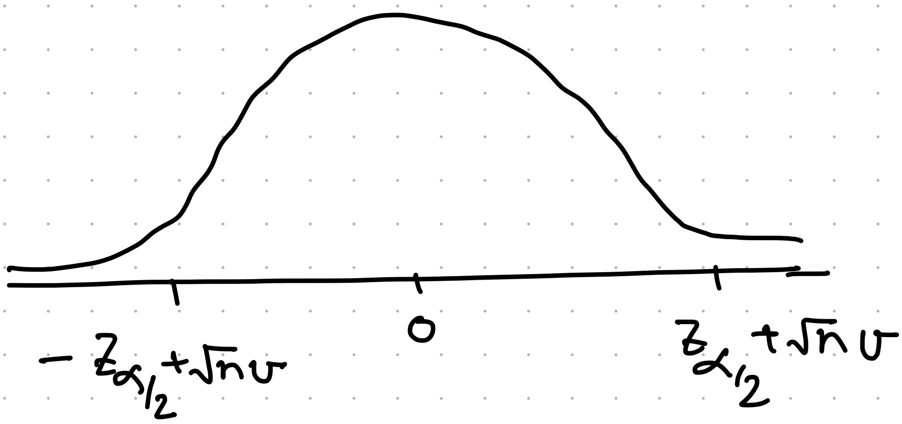
$$\frac{d}{dx}(*) = \frac{d}{dx}\left(F(z_{\alpha/2} + \sqrt{n}v) - F(-z_{\alpha/2} + \sqrt{n}v)\right)$$

$$= f(z_{\alpha/2} + \sqrt{n}v) \cdot \frac{v}{\sqrt{n}}$$

$$- f(-z_{\alpha/2} + \sqrt{n}v)\frac{v}{\sqrt{n}})$$

$$\longrightarrow 0, \quad n \to \infty$$

$$= 1 - \underbrace{P(-z_{\alpha/2} + \sqrt{n}\,v \le Z \le z_{\alpha/2} + \sqrt{n}\,v)}_{F(z_{\alpha/2} + \sqrt{n}\,v) - F(z_{\alpha/2} + \sqrt{n}\,v)}$$



$$\frac{d}{dn}\, g(\mu) > 0$$

$$\text{Ideal},\ g(\mu) = \begin{cases} 0, & z \in \mathbb{R} \\ 1, & z \in R^c \end{cases}$$

$\beta^*_{N+1}(\beta^*_N, X_N, Y_N)$

proof of induction

$$fH:\ \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

$$fS:\ n = 1, \quad 1 = \frac{1(1+1)}{2} = 1.$$

Step: $n \to n+1$

$$\sum_{i=1}^{n+1} i = \underbrace{\sum_{i=1}^{n} i}_{} + n + 1$$

$$= \frac{n(n+1)}{2} + (n+1)$$

$$= \frac{n(n+1)}{2} + \frac{2(n+1)}{2}$$

$$= \frac{(n+2)(n+1)}{2}$$

— x —

$$E = \sum o_k - t_k$$

$$\frac{\partial E}{\partial x^H} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial a_k^H} \cdot \frac{\partial a_k^H}{\partial z^H} \cdot \frac{\partial z^H}{\partial x^H}$$

$$o_k = \sigma(\underbrace{w^H x^H + b^H}_{z^H})$$

$$a^H = \sigma(zH).$$

Extra:

$$Y = AX$$

$Y \in \mathbb{R}^m$ $\quad$ $A \in \mathbb{R}^{m \times n}$

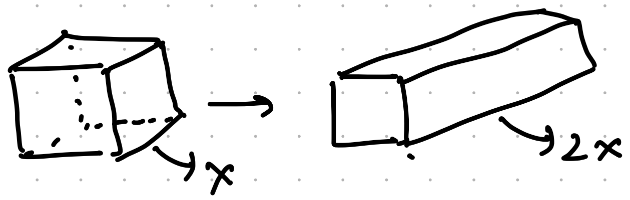
$X \in \mathbb{R}^n$

vol V of point cloud.

$A \in \mathbb{R}^{m \times n}$

vol W

$W = \det A \cdot V$

$V =$ 

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \det A = 2$$

v is an EV with Eigen value λ

$\Leftrightarrow A v = \lambda v$

$v \in \mathbb{R}^m, \ \lambda \in \mathbb{R}$

$[v_{1, \dots, m}] = V$ orthonormal.

$$A = V^T \Sigma V \cdot \quad ,$$

$$\uparrow$$

$$\begin{bmatrix} \lambda_1 & \cdots & & 0 \\ \vdots & \lambda_2 & & \vdots \\ & & \ddots & \\ 0 & \cdots & & \lambda_n \end{bmatrix}$$

$$\det A = \prod_{i=1}^{n} \lambda_i$$

$$\det A = \det (V^T \Sigma V)$$

$$= \det V^T \cdot \det \Sigma \cdot \det V$$

$$= (\det V)^{-1} \cdot \det \Sigma \cdot \det V$$

$$= \det \Sigma$$

$$= \prod_{i=1}^{n} \lambda_i$$

① Exercise 5.3:

$$\text{LRT} = \begin{cases} \exp(n(\theta - X_{(1)})) , & X_{(1)} > \theta \\ 1, & X_{(1)} \leq \theta. \end{cases}$$

② $\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

By calculation, $\hat{\beta} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$

$$\hat{\beta}_{Ridge} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

Which one is good $\hat{\beta}$ or $\hat{\beta}_{Ridge}$??

15 December:

Problem 1: $B_N^* = (X_N^T X_N)^{-1} X_N^T Y_N$

$$X_N = \begin{bmatrix} 1 & x_{11} \cdots x_{1,n} \\ \vdots & \vdots \\ 1 & x_{m,1} \cdots x_{m,n} \end{bmatrix}$$

$$\beta^*_{N+1} = (X^T_{N+1} X_{N+1})^{-1} X^T_{N+1} Y_N$$

$$\mathbb{X}_{N+1} = \begin{bmatrix} X_N \\ X_{N+1} \end{bmatrix} \quad \mathbb{Y}_{N+1} = \begin{bmatrix} Y_N \\ Y_{N+1} \end{bmatrix}$$

$$(X^T_{N+1} X_{N+1})^{-1} = (X_N^T X_N + X^T_{N+1} X_{N+1})^{-1}$$

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + u^T A^{-1} v}$$

[Sherman Morrwood Lobrry]

$$(\mathbb{X}^T_{N+1} \mathbb{X}_{N+1})^{-1} = (\mathbb{X}^T_N \mathbb{X}_N + x^T_{N+1} x_{N+1})^{-1}$$

$$= \underbrace{(\mathbb{X}^T_N \mathbb{X}_N)^{-1}}_{A} - \alpha (\mathbb{X}^T_N \mathbb{X}_N)^{-1} x^T_{N+1} x_{N+1} (\mathbb{X}^T_N \mathbb{X}_N)^{-1}$$

$$\alpha = \left(1 + x_{N+1} (\underline{X}_N^T \underline{X}_N)^{-1} x_{N+1}^T\right)$$

$$\beta^*_{N+1} = (\underline{X}_{N+1}^T \underline{X}_{N+1})^{-1} \underline{X}_{N+1}^T \underline{Y}_{N+1}$$

$$= (\underline{X}_N^T \underline{X})^{-1} (\underline{X}_N \underline{Y}_N + x_{N+1}^T Y_{N+1})$$

$$- \alpha (\underline{X}_N^T \underline{X}_N)^{-1} x_{N+1}^T x_{N+1} (\underline{X}_N^T X)^{-1}$$

$$(\underline{X}_N^T \underline{Y}_N + x_{N+1}^T Y_{N+1})$$

$$K := \alpha (x_N^T x_N)^{-1} x_{N+1}^T$$

$$\beta^*_{N+1} = \beta^*_N + K (Y_{N+1} - x_{N+1} \beta^*_N)$$

$$\begin{bmatrix} \underline{X}_N \\ x_{N+1} \end{bmatrix}^T \begin{bmatrix} \underline{X}_N \\ x_{N+1} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{X}_N^T & x_{N+1}^T \end{bmatrix} \begin{bmatrix} \underline{X}_N \\ x_{N+1} \end{bmatrix}$$

$$= \underline{X}_N^T \underline{X}_N + x_{N+1}^T x_{N+1}.$$

Problem 2:

$$E = \frac{1}{2} \sum_{k \in N_0} (o_k - t_k)^2$$

$$= \frac{1}{2} \| o - k \|_2^2$$

sigmoid function

$$O = \sigma(\omega^H x^H + k^H)$$

$$z^H = \omega^H x^H + k^H$$

$$\Rightarrow O = \sigma(z^H)$$

$$\frac{\partial E}{\partial \omega_{ij}^H} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial z^H} \cdot \frac{\partial z^H}{\partial \omega_{ij}^H}$$

'$\circ$' instead of '$\cdot$'

↳ Hadamard Production

$$a \circ b = \begin{matrix} a_1 \cdot b_1 \\ a_2 \cdot b_2 \\ \vdots \\ a_n \cdot b_n \end{matrix}$$

$$\frac{\partial E}{\partial o_k} = \frac{1}{2} \cdot 2 \cdot 1 (o_k - t_k) = o_k - t_k$$

$$\nabla_o E = \begin{pmatrix} \frac{\partial E}{\partial o_1} \\ \\ \frac{\partial E}{\partial o_{N_o}} \end{pmatrix} = o - k$$

$$\boxed{\frac{\partial E}{\partial O} = o - t}$$

$$O = \sigma(z^H)$$

$$\frac{\partial o_k}{\partial z_k^H} = \sigma'(z_k^H)$$

$$= \sigma(z_k^H) \cdot (1 - \sigma(z_k^H))$$

$$\frac{\partial O}{\partial z} = \sigma(z^H) \circ (1 - \sigma(z^H))$$

$$\frac{\partial o_k}{\partial z_j^H} = \begin{cases} 0 & j \neq k \\ \sigma'(z_k^H) & j = k \end{cases}$$

$$O \in \mathbb{R}_1^{N_o}, \quad z^H \in \mathbb{R}^{N_o}$$

$$\begin{matrix} \frac{\partial o_1}{\partial z_1^H} & \cdots & \frac{\partial o_1}{\partial z_{N_o}^H} \\ \frac{\partial o_2}{\partial z_1^H} & \cdots & \frac{\partial o_2}{\partial z_{N_o}^H} \\ & \vdots & \\ \frac{\partial o_{N_o}}{\partial z_1^H} & \cdots & \frac{\partial o_{N_o}}{\partial z_{N_o}^H} \end{matrix}$$

$$z^H = w^H x^H + b^H$$

$$\frac{\partial z_k^H}{\partial w_{ij}^H} = \frac{\partial}{\partial w_{ij}^H} \sum_{l=1}^{N_H} w_{kl}^H \cdot x_l^H + b_k^H$$

$$= \begin{cases} x_j & \text{if } k = 1 \\ 0 & k \neq 1 \end{cases}$$

$$\frac{\partial z}{\partial w_{ij}^H} = x_j^H \cdot e_i \; ; \; e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow \text{ith row}$$

$$\frac{\partial E}{\partial w_{ij}^H} = (o - t) \circ \sigma(z^H) \circ (1 - \sigma(z^H)) \circ e_i \circ x_j^H$$

* Deep Learning: An Introduction for Applied Mathematicians

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

eigen value,
vector
space

$$\det(A - \lambda I) = (\lambda - 8)(\lambda + 1)(\lambda + 1)$$

$$\therefore \lambda_1 = 8, \; \lambda_2 = \lambda_3 = -1$$

eigen vector, $X = z$, $Y = \frac{1}{2} z$ $\Rightarrow$ eigen space $\begin{bmatrix} z \\ \frac{1}{2} z \\ z \end{bmatrix}$

if $z = 1$, $\begin{bmatrix} 1 \\ \frac{1}{2} \\ 1 \end{bmatrix}$

eigen vecton, $\lambda = \lambda_2 = \lambda_3 = -1$

$$X + \frac{1}{2}Y = Z$$

$$\begin{bmatrix} X \\ Y \\ X + \frac{1}{2}Y \end{bmatrix} = X \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + Y \begin{bmatrix} 0 \\ 1 \\ 1/2 \end{bmatrix}$$