Time Series & Supervised learning

$$\{x_i, y_i\}_{i=1}^{N} \rightarrow \frac{1}{N}\sum_{i=1}^{N} \ell(f(x_i), y_i)$$

$$\approx E_{XY}\left[\ell(f(X), Y)\right]$$

$x_i, y_i \sim P_{XY}$

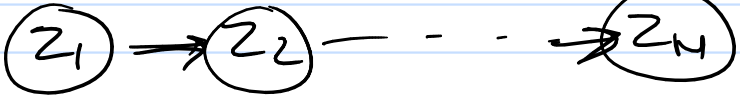
X_i, Y_i

Relation between x_i, y_i & x_j, y_j

$$P(X_i = x_i, Y_i = y_i, X_j = x_j, Y_j = y_j)$$

$$\mathrm{Corr}(X_i, X_j),$$

$x_1\, y_1 \quad x_2\, y_2 \quad \cdots \quad x_N,\, y_N$ order is important.



$$P(z_1 \ldots z_N) = \prod_{i=1}^{N} P(z_i)$$

1 $\Rightarrow P_\theta(z_1, \ldots, z_N)$

3 $\Rightarrow$
$$= \prod_{i=2}^{N} P(z_i \mid z_{i-1}) \cdot P(z_1)$$

2 $\Rightarrow$
$$= P(z_1) \cdot P(z_2 \mid z_1) \cdot \boxed{P(z_3 \mid z_1, z_2)} \cdots P(z_N \mid z_1 \ldots z_{N-1})$$

Previous slide : $\underline{P_\theta(Z_1 \dots Z_N)} \rightarrow$ MLE / EM $\rightarrow \hat{\theta} \rightarrow P(Z_N | Z_{N-1})$

"Simpler" approach : focus on "moments" $E[Z_i]$, $Cov(Z_i, Z_j)$, ...

Weakly stationary seq. of RVs

$Z_1 \dots Z_N$

$\rightarrow E[Z_t] = \mu$ is constant

Covariance $(Z_{t+\tau}, Z_t) = \gamma_\tau < \infty$

Auto covariance function

$$\mathrm{Var}(Z_t) = \gamma_0$$

Auto Correlation $\rho_\tau = \dfrac{\gamma_\tau}{\gamma_0}$

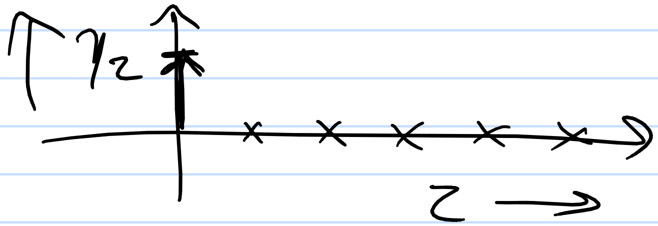
$$\mathrm{Corr}(Z_1, Z_4) = \frac{E[(Z_1-\mu)(Z_4-\mu)]}{\sqrt{\mathrm{Var}(Z_1)\,\mathrm{Var}(Z_4)}}$$

eg 1: $E[Z_t] = 0$ $\quad \mathrm{Var}(Z_t) = 1.$ $\quad \gamma_\tau = 0$ for $\tau = 1, 2, \dots \infty$

$= \sigma^2$

iid

white noise



eg: $\{Z_t\}$ white noise.

$$\left\{\begin{array}{l} S_0 = 0 \\ S_t = Z_1 + \cdots + Z_t. \end{array}\right\} \quad \{S_t\} \text{ Random walk.}$$

$$\text{Covariance}(S_t, S_{t+\tau}) = t\sigma^2.$$

Stationarity : allows us to "average".

$\therefore$ estimate γ_τ

$\mu = \frac{1}{N}\sum_{i=1}^{N} z_i$

$\left.\begin{array}{l}\underline{AR(p)} \\ MA(q)\end{array}\right\}$ Examples of linear TS models.

$AR(1)$: $W_t = \phi W_{t-1} + \varepsilon_t$. $\{\varepsilon_t\}$: white noise

$E[\varepsilon_t] = 0$

$Var(\varepsilon_t) = \sigma^2$

$\cdot 4$

$\longrightarrow \varepsilon_t \perp W_{t-1}, W_{t-2}, \ldots$

Then $\boxed{P(W_1 \ldots W_N) = \prod_{t=2}^{N} P(W_t \mid W_{t-1}) \cdot P(W_1)}$

Seq is called Markovian

$$W_t = \varepsilon_t + \phi(\phi W_{t-2} + \varepsilon_{t-1})$$

$$= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 W_{t-2}$$

$$\vdots$$

$$W_t = \underline{\underline{\varepsilon_t}} + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots \phi^t W_0 \equiv \varepsilon_0$$

$$\boxed{|\phi| < 1}$$

$$\boxed{E[W_t] = 0}$$

$$\gamma_\tau = Cov(W_t, W_{t-\tau}) = \frac{\sigma^2 \phi^{\tau}}{1-\phi^2}$$

MA(q)

$$W_t = \theta_0 \underline{\varepsilon_t} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \cdot \varepsilon_{t-q}.$$

$$W_{t-1} = \theta_0 \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \cdots$$

$\{\varepsilon_t\}$ white noise.

$$+ \theta_q \underline{\varepsilon_{t-q-1}}$$

$$E[W_t] = 0$$

$$\gamma_\tau = \sigma^2 \sum_{k=0}^{q-\tau} \theta_k \theta_{k+\tau} \quad \text{for } 0 < \tau \leq q.$$

$\left.\begin{array}{l}\text{trends}\\ \text{Seasonality}\end{array}\right\}$ predictable patterns in the data.

$$\{Z_t\} \longrightarrow \big(T(t) + S(t)\big)$$

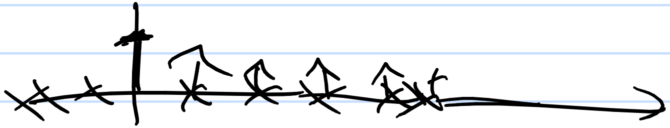
$$\underline{\underline{W_t = \{Z_t\} \text{ minus}}}\ \big(T(t) + S(t)\big)$$

$\longrightarrow$ $\left.\begin{array}{l}\text{AR}\\ \text{MA}\end{array}\right]$

ARMA(p, q) $\quad p = 2, q = 2$

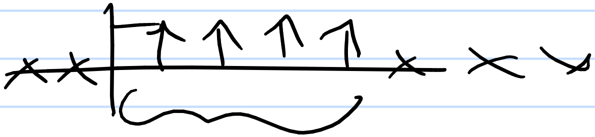
linear. $W_t = \underline{\phi_1} W_{t-1} + \underline{\phi_2} W_{t-2} + \varepsilon_t + \underline{\theta_1} \varepsilon_{t-1} + \underline{\theta_2} \varepsilon_{t-2}$

AR γ_τ



$\{\varepsilon_t\}$ $\quad \sigma^2$

MA γ_τ



$0 < \tau \leq q$

How to estimate $\underline{\phi_1, \phi_2, \theta_1, \theta_2, \sigma^2}$.

<u>Yule-Walker equations?</u>

$$\underbrace{z_1\, z_2 \cdot z_3\, z_4\, z_5}\quad z_6 \quad \cdots \quad z_{100}$$

$$z_6 = f(z_1, \ldots z_5)$$

$$z_7 = f(z_2 \ldots z_6)$$

$$\vdots$$

$$\underbrace{\mathrm{loss}(\beta)}_{\sum_i (y_i - \beta^T x_i)^2} \approx \max \left[\log P_\beta(Y|X) \right] \quad \beta \text{ fixed}$$

$$\| \; + C \cdot \|w\|_2 \approx \left[\; \| \; + \log \mathrm{Prior}(\beta) \right] \quad \beta \text{ RV}$$

$$\underbrace{X_1 \ldots X_p} \sim \boxed{P(X_1 \ldots X_p)} \qquad \{x_i\}_{i=1}^{N} \sim P(\quad) \qquad x \in \mathbb{R}^p$$

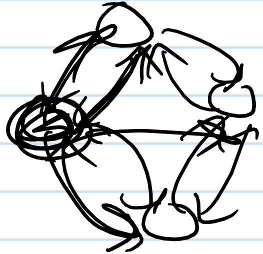
$p = 3$

$$\boxed{\begin{array}{ll} P(X_2 \mid X_1, X_3) & P(X_1 \mid X_2, X_3) \\ P(X_3 \mid X_1, X_2) & \end{array}}$$

x_1

x_2 $\;=\; \boxed{P_\theta(X_2 = x_2 \mid X_1 = x_1)} \rightarrow \equiv$ Markov Chain.

x_3 $\qquad P(X_i = x_i \mid X_{i-1} = x_{i-1})$



$\vdots$

x_N

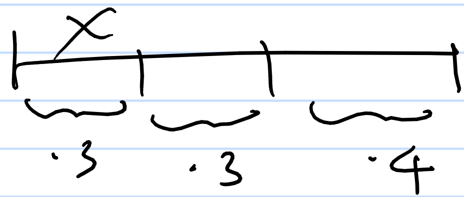
$EM \longrightarrow$ lec 10

K-means.

$\uparrow \; \uparrow$

pmf	2	3	4
	.3	.3	.4

$U[0,1]$



Tree methods. $\longrightarrow$ Gradient boosting.

FSAM. $\longrightarrow$ Adaboost.

$$\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \left(y_i - x_i^T \beta \right)^2$$

L9: Variable importance.

SVM.

L10: $K(x, x') = \phi(x)^T \phi(x') = x^T x'$

γ

ARs

L11: KMeans vs K-medoids

PComponents:

Spectral clustering: