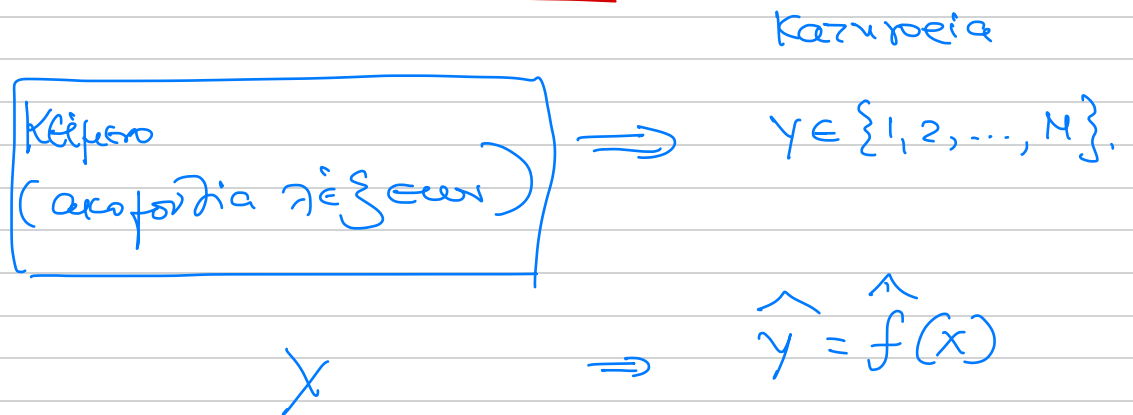


16-5-2025

Document Classification



→ 1) Bag of Words

2) Recurrent Neural Networks (χρονότατες εγγραφές)

1) L όριο λέξεων $\Rightarrow W_1, W_2, \dots, W_L$
 L πρώτες λέξεις του κειμένου

Αναπαράσταση λέξεων.

Dictionary $D = \{ \text{Σύνολο από } M \text{ δυνατά λέξεις (ενιζευμένες)} \}$

Κατασκευή : Σύνολο από προαποφασισμένα κείμενα οι M πιο συχνές λέξεις έχουν αναπαραχθεί

$$D = \{w_1, w_2, \dots, w_M\}$$

$$w_1 = e_1, w_2 = e_2, \dots, w_L = e_L$$

$e_1, \dots, e_M$ μοναδιαία διανύσματα $\in \mathbb{R}^M$

κάθε γέφυρα του κερμένου που βρίσκεται στο D παρουσιάζει από ένα διάνυσμα e_j

κερέο L γέφυρες

$$\begin{matrix} 1 \\ 2 \\ \vdots \\ L \end{matrix} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{matrix} M \\ \vdots \\ 0 \end{matrix} \leftarrow w_1$$

$$\begin{matrix} \vdots \\ 0 \end{matrix} \leftarrow w_L$$

κερέο $\Rightarrow X_1, \dots, X_L = \underline{\underline{X}}$

$$\left(\begin{matrix} \updownarrow & \dots & \updownarrow \\ w_1 & \dots & w_L \end{matrix} \right)$$

Γενικότερο Ερώτημα

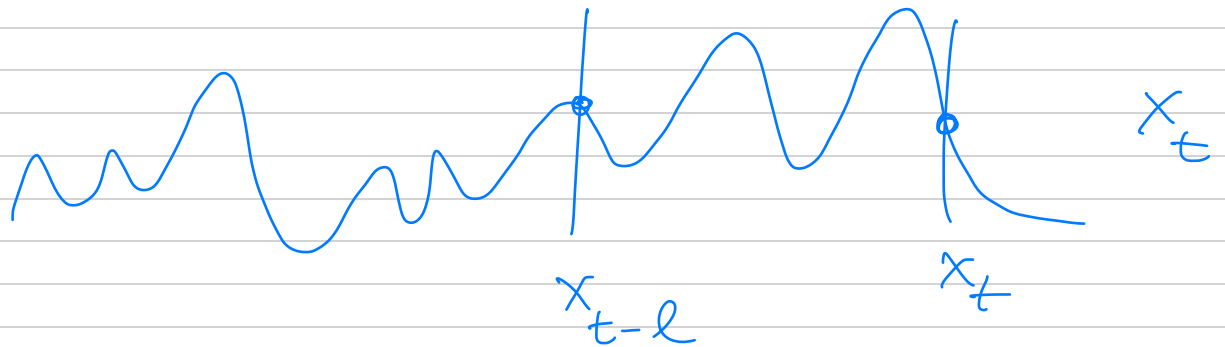
Ακατόντια γέφυρες κ' όχι αλλά μόνο

Εφαρμογή κυρίως σε χρονοσειρές

$\{X_1, X_2, \dots\}$ ακατόντια χρονικά μεταβλητών X_t
 t : χρόνος

Αυτοσυσχέτιση

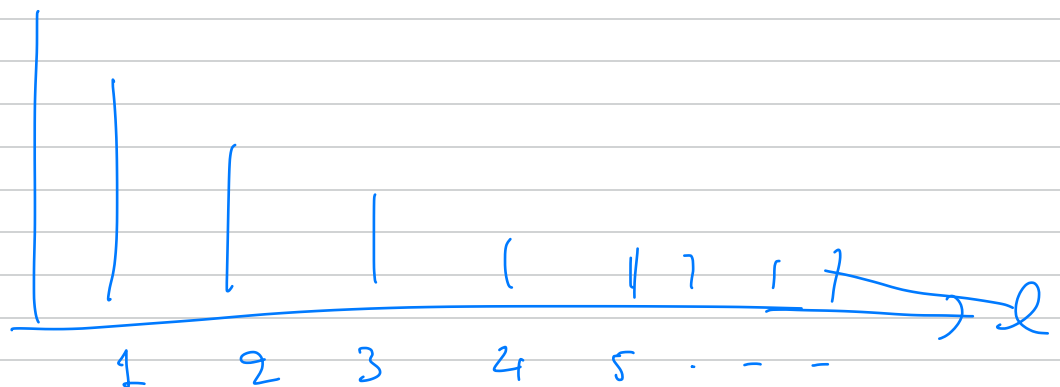
$$r_{t,l} = f(X_t, X_{t-l})$$



Qu $r_{t,l}$ ανέξ. ζωτ $\Rightarrow$ ορίζουν

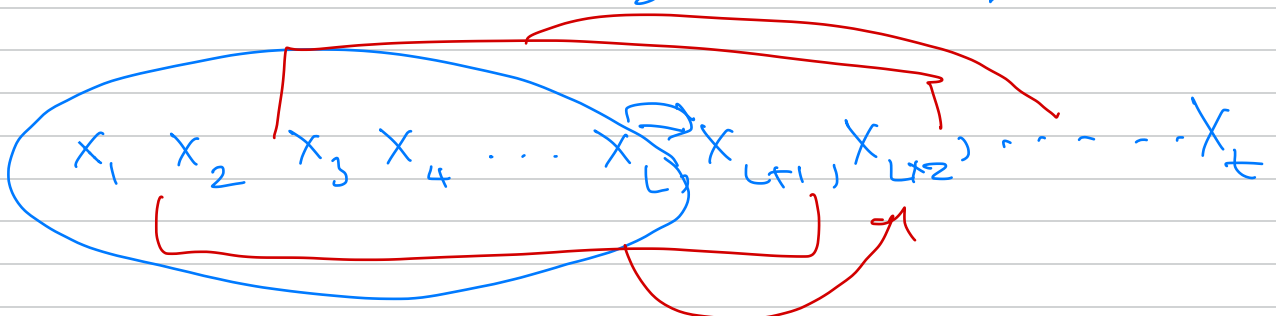
r_l : ορίζουν αυτοσυσχέτισης (autocorrelation function)

Συνάρτ. r_l φθίνει ως προς l .



Autoregressive (AR) model τάξης L

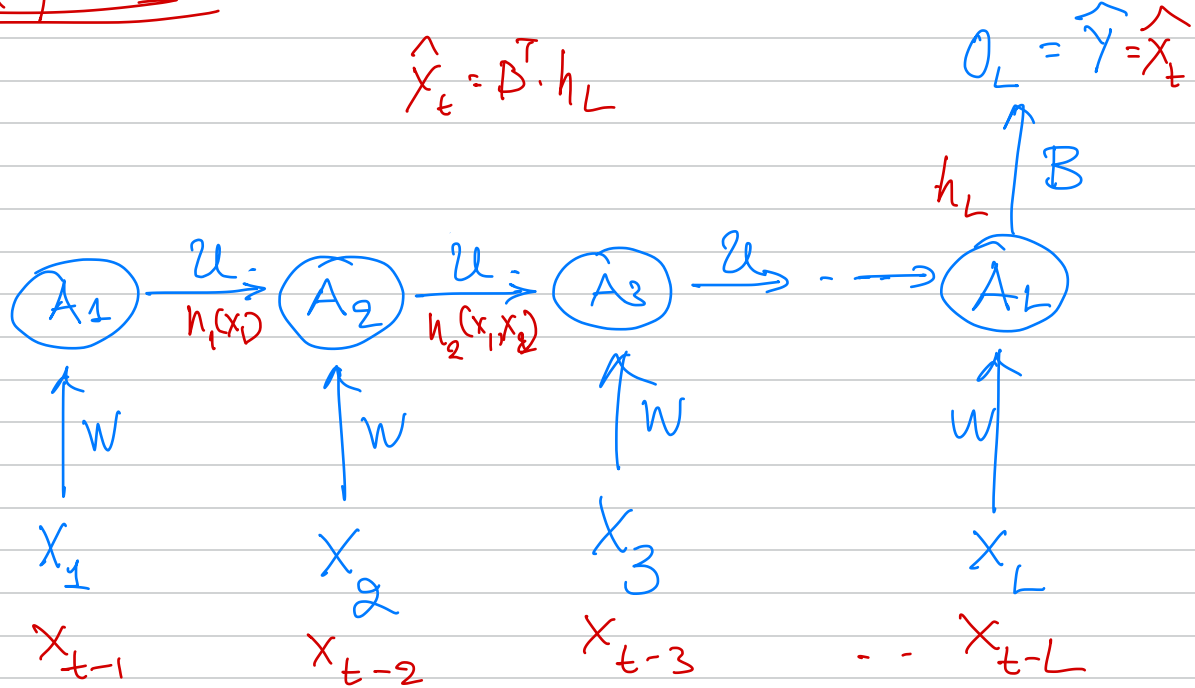
$$X_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} + \dots + b_L X_{t-L}$$



Recurrent Neural Network { "neural network με αυτοσυνεκτιμήσεις" }

Αναδρομική

$$\hat{X}_t = B^T \cdot h_L$$



$$h_1 = g(w^T X_1)$$

$$h_2 = g(u^T \cdot h_1 + w^T \cdot X_2)$$

$$h_k = g(u^T \cdot h_{k-1} + w^T \cdot X_k) \quad \text{αναδρομική (recurrence) σχέση}$$

Εφαρμογή : Πρόβλεψη όγκου συναλλαγών

Dataset. $X_t = \begin{pmatrix} v_t \\ r_t \\ z_t \end{pmatrix}$, $v_t = \log(\text{trading volume})$
 $r_t = \log(\text{DowJones return})$
 $z_t = \log(\text{volatility})$

Trading Volume : όγκος συναλλαγών (average 11 ημερών)

DowJones return (απόδοση)

Εάν θεωρήσουμε τιμή S_t την ημέρα t .

% Απόδοση σχετικά με τον προηγούμενο : $\frac{S_t - S_{t-1}}{S_{t-1}} = R_t$

$$R_t = \frac{S_t}{S_{t-1}} - 1 \Rightarrow \frac{S_t}{S_{t-1}} = \underbrace{(R_t + 1)}_{W_t} \geq 0$$

$$r_t = \log(W_t) \approx \log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1})$$



αριθμός τυχαίου $B(t)$ κίνηση Brown

$$W_t = e^{B(t)} \Rightarrow \text{Geometric Brownian motion}$$

$$B(t) \sim \mathcal{N}(0, \sigma^2 t)$$

$\downarrow$
volatility

