Ashutosh Kumar Singh 19CS30008

8. $\hat{z}_{t+1} = \theta_1 z_t + \dots + \theta_M z_{t-M+1}$, $t = M, M+1, \dots, 100$.

(a) We can use least squares to choose the model parameters $\theta_1, \theta_2, \dots, \theta_M$, based on the observed data $z_1, z_2, \dots, z_{100}$ by minimizing the sum of squares of the prediction errors $z_t - \hat{z}_t$ over $t = M+1, \dots, 100$

i.e., $(z_{M+1} - \hat{z}_{M+1})^2 + \dots + (z_{100} - \hat{z}_{100})^2$.

Note that we start the predictions at $t = M+1$, since any prediction depends on the previous M values.

So, this can be put into the general linear in parameters model form by taking:-

$$y^{(i)} = z_{M+i}, \quad x^{(i)} = (z_{M+i-1}, \dots, z_i),$$
$$i = 1, \dots, 100-M$$

We have $100-M$ examples, and each feature vector has M features.

The equations we have in hand are:-

$$z_{M+1} = \theta_1 z_M + \theta_2 z_{M-1} + \dots + \theta_M z_1$$
$$z_{M+2} = \theta_1 z_{M+1} + \theta_2 z_M + \dots + \theta_M z_2$$
$$\vdots$$
$$z_{100} = \theta_1 z_{99} + \theta_2 z_{98} + \dots + \theta_M z_{100-M}$$

This can be modeled as a least squares problem $A\theta = b$.

(b) For the least squares formulation, the matrix A is:-

$$A = \begin{bmatrix} z_M & z_{M-1} & \cdots\cdots & z_1 \\ z_{M+1} & z_M & \cdots\cdots & z_2 \\ \vdots & \vdots & & \vdots \\ z_{99} & z_{98} & \cdots\cdots & z_{100-M} \end{bmatrix}$$

$A \in R^{(100-M) \times M}$

The vector b is :-

$$b = \begin{bmatrix} z_{M+1} \\ z_{M+2} \\ \vdots \\ z_{100} \end{bmatrix}$$

$b \in R^{(100-M)}$

The vector θ will be:-

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{bmatrix}$$

$\theta \in R^M$

(c) Every time we move one term ahead in the time-series model, the previous row is shifted by one position to the ~~left~~ right, so the ~~left~~ rightmost entry is removed and one new entry is added to the ~~right~~ left.

i^{th} row $= [z_{i+M-1} \quad z_{i+M-2} \quad \cdots \quad z_i]$

$(i+1)^{th}$ row $= [z_{i+M} \quad z_{i+M-1} \quad \cdots \quad z_{i+1}]$

This is the special structure we can observe in A. Also, note that diagonals of the matrix A contain the same values.

$$\begin{bmatrix} z_M & z_{M-1} & z_{M-2} & \cdots & z_1 \\ z_{M+1} & z_M & z_{M-1} & \cdots & z_2 \\ z_{M+2} & z_{M+1} & z_M & \cdots & z_3 \\ \vdots & & & & \\ z_{99} & z_{98} & z_{97} & \cdots & z_{100-M} \end{bmatrix}$$

(d) No two rows or columns of the matrix will be the same, assuming that the points do not obey some specific special structure.

Hence, since there are $(100-M)$ rows and M columns, we can say that :-

$$\text{rank}(A) \leq \min(M, 100-M)$$