$\boxed{\text{Q3.1}}$

$$h_i = z_i(1-z_i) = z_i - z_i^2$$

we know:

$$z_i = \sum_j W_{ij} x_j + b_i$$

$$z_i^2 = \left(\sum_j W_{ij} x_j + b_i\right)^2 = \left(\sum_j W_{ij} x_j\right)^2 + 2\sum W_{ij} x_j \cdot b_i + b_i^2$$

$$\left(\sum W_{ij} x_j\right)^2 = \sum (W_{ij} x_j)^2 + 2\sum_j \sum_{m \neq j} (W_{ij} x_j)(W_{im} x_m)$$

therefore:

$$h_i = \sum_{j=1}^{D} W_{ij} x_j + b_i - \sum_{j=1}^{D} W_{ij}^2 x_j^2 - 2\sum_{j=1}^{D} \sum_{m \neq j} (W_{ij} x_j) \cdot (W_{im} x_m) - 2\sum_{j=1}^{D} W_{ij} x_j \cdot b_i - b_i^2$$

feature transformation

ϕ_K:

$K = 1 \rightarrow \quad \phi_1(x) = 1$

$K = 2$ to $K = D+1 \rightarrow \quad \phi_{i+1} = x_i \quad$ for $i = 1, \dots, D$

$K = D+2$ to $K = 2D+1 \rightarrow \quad \phi_{i+1+D} = x_i^2 \quad$ for $i = 1, \dots, D$

$K = 2D+2$ to $K = \frac{(D+1)(D+2)}{2} \rightarrow \quad \phi_K = x_i \cdot x_j \quad$ for $i = 1, \dots, D \quad j = i+1, \dots, D$

$$\phi(x) = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_D \\ x_1^2 \\ \vdots \\ x_D^2 \\ x_1 x_2 \\ \vdots \\ x_1 x_D \\ x_2 x_3 \\ \vdots \\ x_2 x_D \\ \vdots \\ x_{D-1} x_D \end{pmatrix}$$

hi can be re-written as:

$$h_i = b_i - b_i^2 + (1-2b_i)\sum_{j=1}^{D} W_{ij} x_j - \sum_{j=1}^{D} W_{ij}^2 x_j^2 - 2\sum_{j=1}^{D}\sum_{m \neq j} (W_{ij} \cdot x_j)(W_{im} x_m)$$

$$= A_\Theta \cdot \Phi \qquad \text{wher } A_\Theta \text{ is:}$$

$$A_\Theta = \begin{pmatrix} b_1 - b_1^2 & (1-2b_1)W_{11} \dots (1-2b_1)W_{1D} & -W_{11}^2 \dots - W_{1D}^2 & -2W_{11}W_{12} \dots -2W_{11}W_{1D} & -2W_{12}W_{13} \dots -2W_{12}W_{1D} & \dots -2W_{1D-1} W_{D} \\ \vdots & & & & & \\ b_K - b_K^2 & (1-b_K)W_{K1} \dots (1-2b_K)W_{KD} & -W_{K1}^2 \dots -W_{KD}^2 & -2W_{K1}W_{K2} \dots -2W_{K1}W_{KD} & -2W_{K2}W_{K3} \dots -2W_{K3}W_{KD} & \dots -2W_{KD-1}W_{D} \end{pmatrix}$$

$\boxed{Q3.2}$

$$\hat{y} = v^T h + v_0 \qquad v^T \in \mathbb{R}^K$$
$$h = A_\Theta \Phi(x) \qquad A_\Theta \in \mathbb{R}^{K \times \frac{(D+1)(D+2)}{2}}$$

$$\hat{y} = v^T h + v_0 = v^T A_\Theta \Phi(x) + v_0 = C_\Theta^T \Phi(x)$$

therefore $C_\Theta^T = v^T A_\Theta + \begin{pmatrix} v_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T$ and $C_\Theta = A_\Theta^T v + \begin{pmatrix} v_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

In terms of the original parameters it is not a linear model since $\Phi(x)$ is not a linear transformation.