

$$g([x^{(n)}]^T W_1)$$

nonlinear function applied to each element of the vector. (element-wise function)

To make life easier, see bias as input with value 1.
 A format used for the rest of this course (hmm, not quite but close enough)

$$X \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} \quad \text{inner product}$$

$$XW + C^T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \quad \text{Transformed features}$$

$$g(XW + C^T) = \max\{XW + C^T, 0\} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}$$

$$g(XW + C^T)W = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \quad \begin{matrix} w = [-1 \\ -2] \end{matrix}$$

Idea is to find W & w

