02/07/2023 Neural Networks

1) Review of SVD

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = U \Sigma V^T$$

Data Matrix

columns of U provide a good model for columns of A

$a^* = U\alpha$

Side note: Linear Operator Definition

$A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2$

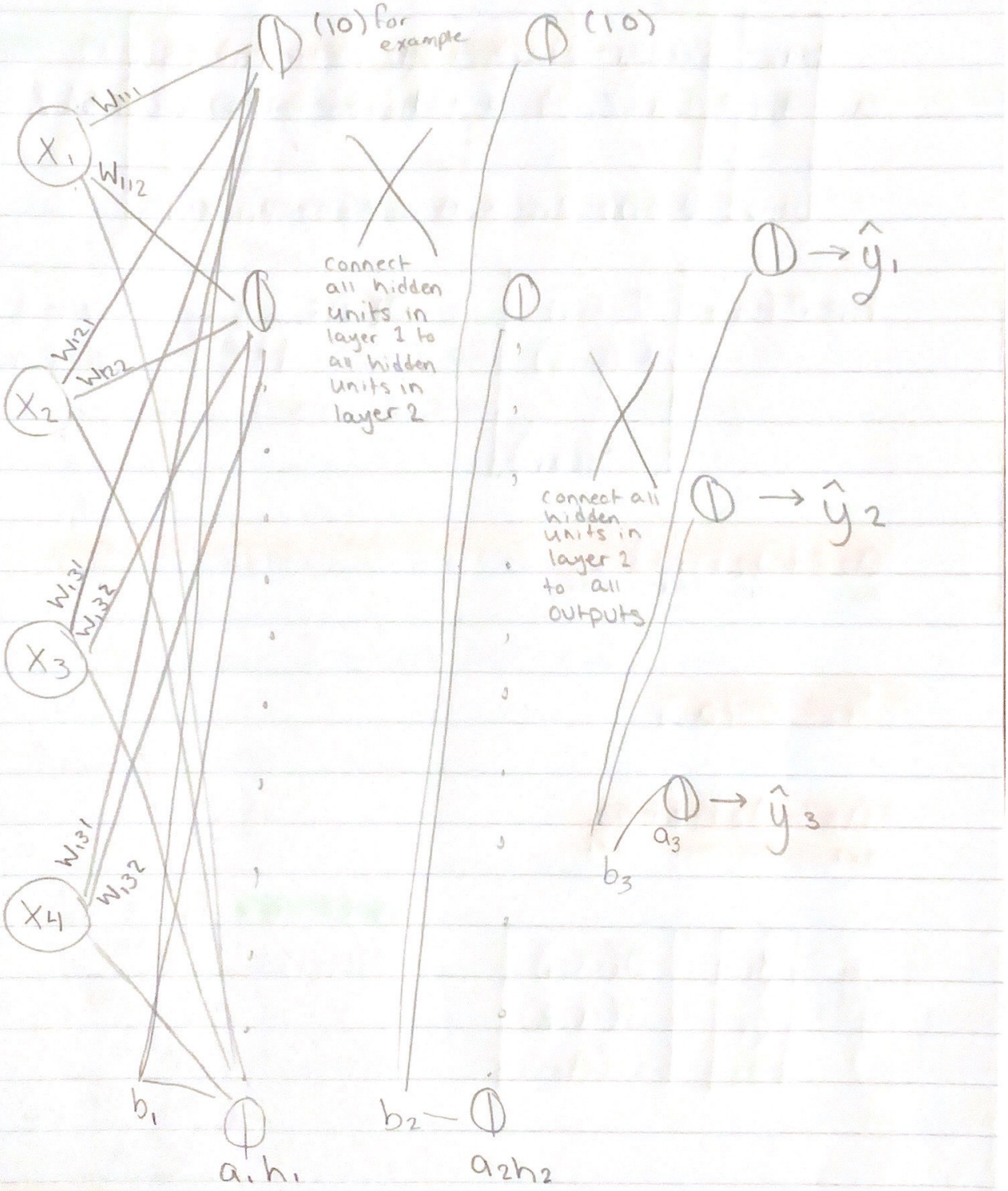
- SVD is linear because W is definitely separated from U and W is the parameters of A; U and W are in product form

2) Neural Networks

- Sklearn's Iris dataset
  - independent variables
    - petal length, petal width, sepal length, sepal width
  - classes (1, 2, 3)
- Multilayer Perceptron / Artificial Neural Networks
  - weights

W_{abc}

where a = hidden layer number
where b = input variable number
where c = hidden unit number



$a_1 = W_1 X$

$10 \times 4 \quad 10 \times 4 \quad 4 \times 1$

$$a_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{110} \end{bmatrix} = \begin{bmatrix} W_{111}X_1 + W_{121}X_2 + W_{131}X_3 + W_{141}X_4 + b_1 \\ W_{112}X_1 + W_{122}X_2 + W_{132}X_3 + W_{142}X_4 + b_1 \\ \vdots \\ W_{1110}X_1 + W_{1210}X_2 + W_{1310}X_3 + W_{1410}X_4 + b_1 \end{bmatrix} = W_1X + b_1$$

$$h_1 = \sigma(a_1) = \begin{bmatrix} \sigma(a_{11}) \\ \sigma(a_{12}) \\ \vdots \\ \sigma(a_{110}) \end{bmatrix} \qquad \sigma(v) = \frac{1}{1+e^{-v}} \text{ (sigmoid function)}$$

$$\underset{10\times1}{a_2} = \underset{10\times10}{W_2}\,\underset{10\times1}{h_1} + \underset{10\times1}{b_2}$$

$$h_2 = \sigma(a_2)$$

$$\underset{3\times1}{a_3} = \underset{3\times10}{W_3}\,\underset{10\times1}{h_2} + \underset{3\times1}{b_3}$$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} o(a_{31}) \\ o(a_{32}) \\ o(a_{33}) \end{bmatrix}$$

Soft max function

$$o(a_{31}) = \frac{e^{a_{31}}}{e^{a_{31}} + e^{a_{32}} + e^{a_{33}}}$$

$$o(a_{32}) = \frac{e^{a_{32}}}{e^{a_{31}} + e^{a_{32}} + e^{a_{33}}}$$

$$o(a_{33}) = \frac{e^{a_{33}}}{e^{a_{31}} + e^{a_{32}} + e^{a_{33}}}$$

parameters: $W_1, W_2, W_3, b_1, b_2, b_3$

W_1: 10×4, W_2: 10×10, W_3: 3×10, b_1: 10, b_2: 10, b_3: 3

$40 + 100 + 30 + 23 = 173$ parameters

a represents the intermediate steps with the weights and bias

h represents the result of putting a in a function, this gives us the non-linearity aspect

$$y = O(W_3(\sigma(W_2(\sigma(W_1 X + b_1)) + b_2)) + b_3)$$

- Other non-linear function options
  - tanh
  - $ReLU(v) = \max(0, v)$