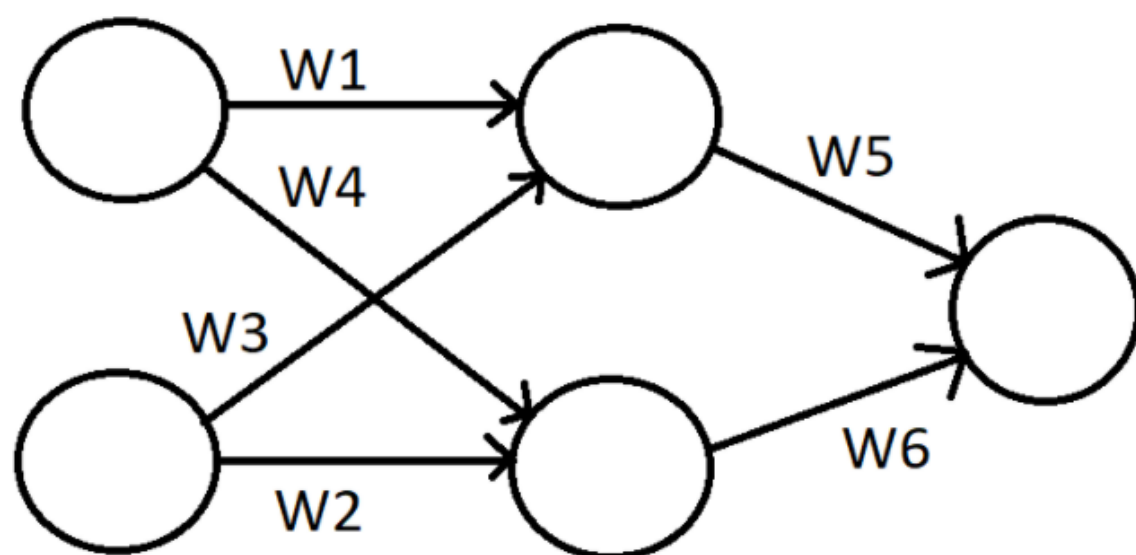


Task

Consider the network



Where w_i represent the weights. The first layer is the input layer, the second layer uses a Sigmoid activation function. The weights have the following input:

W1	1
W2	2
W3	0.5
W4	-1
W5	-2
W6	-0.5

And we have the input:

	Feature 1	Feature 2
Sample 1	1	20
Sample 2	25	3
Sample 3	16	2

FIND:

- Normalize the input data
- Do one forwards pass through the network. You can choose γ , β and α yourself, but the solution uses 0.5, 0.5 and 0.1 respectively. The Network uses is implemented with Batch Normalization.
- The paper suggest computing sample mean and variance for each feature using a large number of training samples, is there an alternative approach to this? If so, can you perform the necessary calculation based on the forward pass? (hint: moving averages)