1 Neural network language model

1. Equation
   h1(x,θ_1, θ_2, b_1) = $\sigma(\theta_1 x1 + \theta_2 x2 + b_1)$
   $h2(x, \theta_3, \theta_4, b_2) = \sigma(\theta_3 x1 + \theta_4 x2 + b_2)$
   $f(x, \theta, b, w_1, w_2, c) = \sigma(w_1 h1(x, \theta, b1) + w_2 h2(x, \theta, b2) + c)$

2. Dimensionality

	input	weight
Dimensionsality	n(n neurons)	n x $\|\|v\|\|(vocabulary)$

2 Sanity Check

1. Hyper-parameters

	Epoch=1	Epoch=1000	Epoch=500	Epoch=400
Learning Rate = 0.01	Fail	Pass	Pass	Pass
Learning Rate = 0.001	Fail	Pass	Pass	Fail
Learning Rate = 0.0001	Fail	Fail	Fail	Fail

2. Q : Why "Start the mathematician instead of "Start the physicist"?
   A : Because the probability of mathematician is bigger

	mathematician	physicist
Probability	tensor(-0.5306)	tensor(-1.6375)

3 Test

1. Hyper-parameters : see sanity check above
   The Learning Rate was set to 0.001 and Epoch was set to 500

2. Q : Discuss whether this would be possible with the bigram ML model from lab 2.
   A : I think it is possible to use bigram because we can still count the similarity between each other words. However, the correctness of the result might be lower than trigram.

3. Embeddings similarity with mathematician

	philosopher	physicist
Similarity	tensor([0.3577]	tensor([0.1659]