1.

Input	HL 1	HL 2	Output

X_1

X_2

y

$\cdot \, (\text{Input} + \text{output}) \cdot \dfrac{2}{3} = \boxed{2 \ \text{neurons per layer}}$

2. Each neuron contributes by evaluating the logistic function to separate between the dark and white areas. The first hidden layer applies decision layers to the shape of the mask and the second layer combines these regions To separate Class 1 (Dark) and class 2 (light) to form complex boundaries.

3. A single layer network is not capable of producing this mask because it is significantly limited to linearly separable data.

- the universal approximation theorem doesn't hold because the 'mask' is not a continuous function.