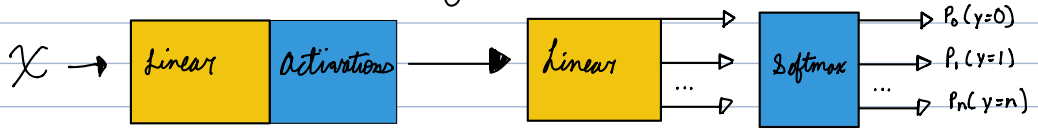
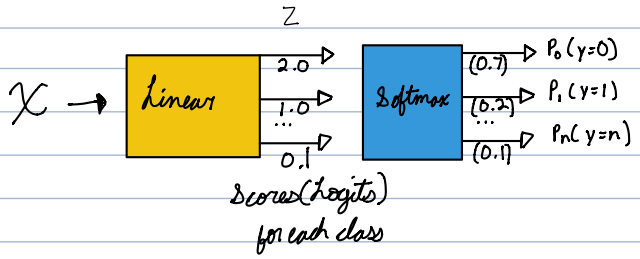
Softmax





Softmax: $\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$ for $j = 1, \dots, K$

The sum of sigmoid always = 1

Cross Entropy: way to quantify the difference between the learned probability and ground truth distribution

$$L = \frac{1}{N} \sum_i \underbrace{D(s(wx_i + b), y_i)}_{D(\hat{Y}, Y)} = -Y \log \hat{Y}$$

The goal for the model when classifying each class is have an one hot encoded problem.

Class	One hot encoding
0	[1, 0, 0, 0, 0]
1	[0, 1, 0, 0, 0]
2	[0, 0, 1, 0, 0]
3	[0, 0, 0, 1, 0]
4	[0, 0, 0, 0, 1]