HW2

Problem 1 For multi-layer perceptron,

$$s^{(l)} = W^{(l)} h^{(l-1)} + b^{(l)},\ h^{(l)} = \sigma(s^{(l)}),\ l = 1, ..., L,$$

where $s = s_L = h_L$ is the top layer logit score for logistic regression, $h_0 = x$ is the input, and $\sigma()$ is element-wise ReLU.

(1) Derive $\partial \log p(y|s)/\partial s$ at the top layer.

(2) Derive the back-propagation algorithm. You can first derive results in terms of elements of matrices and vectors, and then arrange the results in terms of matrices and vectors.

Problem 2 Write Python code to train MLP in Problem 1, using Adam optimizer. The code should allow flexible number of layers and number of nodes in each layer. Apply your code to the classification problem in HW1. You can decide on the details.