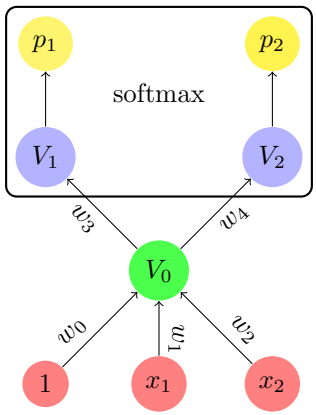
Homework-4 Solutions

Question 1



The above neural network has two inputs. It computes a selection between the two alternatives A, B in terms of two probability outputs. p_1 is the probability that A occurs, and p_2 is the probability that B occurs. The node V_0 is implemented with ReLU. The nodes V_1, V_2 are linear (ADALINE), and they are not connected to a bias. The probabilities p_1, p_2 are computed from the values of V_1, V_2 using softmax.

A.1: Compute the values of all nodes in forward propagation when the network is given the input $\boxed{x_1 = 2,\ x_2 = 7}$, the current weight values are: $\boxed{w_0 = 0,\ w_1 = 0.2,\ w_2 = 0.1,\ w_3 = 0.1,\ w_4 = 1}$, with $\boxed{\text{the desired selection being } \mathbf{A}}$. Use training rate $\boxed{\epsilon = 0.1}$. Your answer should be explicit numeric values for each node.

Answer

$$
\begin{aligned}
V_0 &= w_0 + 2w_1 + 7w_2 = 0.4 + 0.7 = 1.1 \\
V_1 &= w_3 V_0 = 0.11 \\
V_2 &= w_4 V_0 = 1.1 \\
p_1 &= e^{V_1}/(e^{V_1} + e^{V_2}) = 0.27 \qquad (Z = 4.12) \\
p_2 &= e^{V_2}/(e^{V_1} + e^{V_2}) = 0.73
\end{aligned}
$$

A.2: Compute explicit numeric values to the weights after they are changed by back propagation when the network is given the example above.

Please show the values of the temporary "delta" variables.

Answer The values of the temporary delta variables:

$$\delta_1 = 1/2(1 - p_1) = 0.365$$
$$\delta_2 = 1/2(-p_2) = -0.365$$
$$\delta_0 = (w_3\delta_1 + w_4\delta_2)g'(V_0) = (0.1 - 1)0.365(1) = -0.3285$$

$$\text{new } w_0 = w_0 + \epsilon\delta_0(1) = -0.03285$$
$$\text{new } w_1 = w_1 + \epsilon\delta_0 x1 = 0.1343$$
$$\text{new } w_2 = w_2 + \epsilon\delta_0 x2 = -0.12995$$
$$\text{new } w_3 = w_3 + \epsilon\delta_1 V_0 = 0.14015$$
$$\text{new } w_4 = w_4 + \epsilon\delta_2 V_0 = 0.95985$$

Question 2

Consider a deep neural net applied to decide between the following four categories:

cat, tiger, human face, lion

The neural net uses a softmax unit at the output layer. Consider the case where the values fed into the output layer are:

cat	0.5
tiger	0.8
human face	-3
lion	0.6

The softmax converts these values into a probability vector.

1. Compute the probabiity vector.

2. Which outcome is the most likely?

3. Which outcome is the least likely?

4. What is the result of cross-entropy cost function if the target output is lion?

Question 3

The Adam technique for accelerating back propagation was specified in terms of the following parameters: λ, α, ϵ, β_1, β_2. The standard back propagation algorithm was specified in terms of the learning-rate parameter ϵ. Show how to select the parameters of Adam so that the result is as close approximation to standard back propagation as possible.

$\lambda = 0$, $\alpha = \epsilon$, $\epsilon = 0$, $\beta_1 = 0$, $\beta_2 \approx 1$ gives:

$$W_t = W_{t-1} - \epsilon \frac{D_t}{\sum_{i=1}^{t} D_t^2/t}$$

For large values of t we expect the mean at the denominator to be approximately a constant.

The following is incorrect. $\lambda = 0$, $\alpha = \epsilon$, $\epsilon = 0$, $\beta_1 = 0$, $\beta_2 = 0$ gives:

$$W_t = W_{t-1} - \epsilon \frac{D_t}{D_t} = W_{t-1} - \epsilon$$