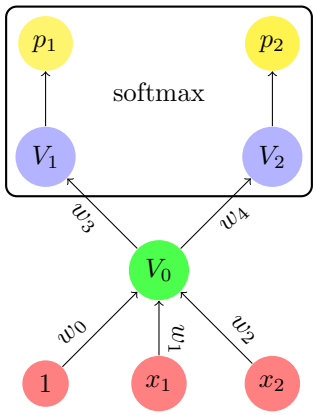
Homework-3 Solutions

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

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

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.

Answer:

$$e^{0.5} = 1.64872, \quad e^{0.8} = 2.22554, \quad e^{-3} = 0.0497871, \quad e^{0.6} = 1.82212, \quad e^{0.5}+e^{0.8}+e^{-3}+e^{0.6} = 5.74617$$

$$p = (0.286925, 0.387309, 0.00866439, 0.317102)$$

2. Which outcome is the most likely?

Answer: tiger

3. Which outcome is the least likely?

Answer: human face

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

Answer:

$$E = \ln(1/0.317102) = 1.14853$$

Question 3

In the table below cases 3,4 are distributions, and cases 1, 2 can be converted into distributions.

case	A	B	C	D
1	1	-2	3	-4
2	1	2	-3	0
3	1	0	0	0
4	1/4	1/4	1/4	1/4

Converting 1 into a probability distribution using softmax:

$$V = \{1, -2, 3, -4\}$$
$$q = \{2.71828, 0.135335, 20.0855, 0.0183156\}$$
$$Z = 22.9575$$
$$p = \{0.118405, 0.00589504, 0.874902, 0.000797807\}$$

Converting 2 into a probability distribution using softmax:

$$V = \{1, 2, -3, 0\}$$
$$q = \{2.71828, 7.38906, 0.0497871, 1\}$$
$$Z = 11.1571$$
$$p = \{0.243636, 0.662272, 0.00446236, 0.0896288\}$$

1. Use cross entropy to determine which distribution among 1,2,3 is most similar to 4. **Show your computations.**

case	A	B	C	D	cross entropy of p_4 with candidate:
1	0.118405	0.00589504	0.874902	0.000797807	5.24224
2	0.243636	0.662272	0.00446236	0.0896288	3.47989
3	1	0	0	0	infinity
4	1/4	1/4	1/4	1/4	2

Answer: 1 / $\boxed{2}$ / 3

2. Use cross entropy to determine which distribution among 1,2,4 is most similar to 3. **Show your computations.**

case	A	B	C	D	cross entropy of p_3 with candidate:
1	0.118405	0.00589504	0.874902	0.000797807	3.0782
2	0.243636	0.662272	0.00446236	0.0896288	2.0372
3	1	0	0	0	0
4	1/4	1/4	1/4	1/4	2

Answer: 1 / 2 / $\boxed{4}$

Question 4

In this question, if you need to compute logarithms use natural basis logarithm (ln).

Consider a deep neural net applied to decide between the following three categories: A, B, C.

1.

1.1 What is the one-hot encoding of the category A?

Answer: (1,0,0)

1.2 What is the one-hot encoding of the category B?

Answer: (0,1,0)

1.3 What is the one-hot encoding of the category C?

Answer: (0,0,1)

2.

Consider the deep learning network output specified by the vector $z = (1.0, 2.0, 3.0)$. Using softmax and the cross entropy criterion is z closer to A, to B, or to C (in their one-hot encoding)?

Answer:

A / B / $\boxed{\text{C}}$

3.

Find a bias value x (it can be either positive or negative) such that the vector $z_x = (1.0 + x, 2.0 + x, 3.0 + x)$ would be considered by the network as closer to A than to B or to C.

Answer.

no such value exist

4.

Find a bias value x (it can be either positive or negative) such that the vector $z_x = (1.0 + x, 2.0 + x, 3.0 + x)$ would be considered by the network as closer to B than to A or to C.

Answer.

No such value exist.

5.

Find a bias value x (it can be either positive or negative) such that the vector $z_x = (1.0 + x, 2.0 + x, 3.0 + x)$ would be considered by the network as closer to C than to A or to B.

Answer.

any value.

Details

The cross entropy of the vector p with the vector q is $H(p, q) = \sum_j p_j \log(1/q_j)$. When p is obtained from one-hot encoding with the hot coordinate being j, the result is: $H(p, q) = \log(1/q_j)$. Therefore $H(p, q)$ is smallest when q_j is largest. Regardless of which value is assigned to x the softmax value associated with $3 + x$ will always be the largest of the three coordinates. To see this observe that after softmax the value of q_j is $\frac{e^{j+x}}{e^{1+x}+e^{2+x}+e^{3+x}}$ for $j = 1, 2, 3$. This reduces to $\frac{e^j}{e^1+e^2+e^3}$ for $j = 1, 2, 3$, which is independent of x.