

Problem 1: Setting up a Neural Network

Suppose we want to train a neural network that takes in an image of size 40×40 and outputs “capybara”, “giraffe”, or “porcupine”. We would like to use one-hot encoding to encode the output.

Sketch the neural network architecture. What is the size of the input layer? What is the size of the output layer?

There are three possible values for the correct values for the output layer. What are they? (Review one-hot encoding).

Problem 2: Cost functions

For a **single** training example (an image of a capybara) x with expected output y , the we can formulate least-squares cost as

$$\sum_{k=1}^3 (net(x)_k - y_k)^2$$

Write down the cost function for the entire training set

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(M)}, y^{(M)})\}$$

Problem 3: Likelihood

Consider a network with a softmax layer at the top, i.e., one that outputs the three probabilities $p_1 = \text{net}_W(x)_1, p_2 = \text{net}_W(x)_2, p_3 = \text{net}_W(x)_3$. Suppose the weights of the network are not changing, and are W . If we interpret p_1, p_2, p_3 as the probabilities that the network assigns to {"capybara", "giraffe", "porcupine"}.

Suppose the correct output for the input x is "capybara." What is the likelihood for that input? What is the log-likelihood for that input?

Now, use the trick from last lecture to write a formula for the log-likelihood of the input x by summing over the components of y .

Write down the log-likelihood of

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), (x^{(4)}, y^{(4)})\}$$