9.641 Neural Networks
Problem Set 2: Perceptrons and linear networks

(Due before class on Thursday, Feb. 17)

You are expected to turn in any code you might have used to answer a question.

1. About linear perceptrons

    (a) We first consider the toy problems from Fig. 1 for which we want to separate filled from empty circles. Draw by hand (and submit the resulting figure) a possible separation surface for each of the examples and name their corresponding logical operation.

    Hint: Recall that in 2D, the decision boundary for a simple perceptron is a line.

    For each problem, propose a perceptron architecture (*i.e.* a network diagram and its associated weights). Although there are many possible solutions we ask for a single one.

    (b) Show that a simple perceptron cannot solve the XOR problem from Fig. 2.

    Hint: The general equation of a perceptron with two inputs is given by:

    $$w_0 + w_1 X_1 + w_2 X_2 = 0 \tag{1}$$

    Assume that there exists a solution and find a system of equations that the perceptron coefficients must satisfy. Show that this leads to a contradiction.

    (c) Show that a two-layer perceptron can solve the XOR problem. Draw a possible solution to the problem on Fig. 2. Submit a figure and a network diagram (with associated weights).

    Hint: You can solve the problem by combining the output of two simple perceptrons with an AND function.

    (d) The previous example illustrates how adding more layers to a network allows for more complex decision boundaries. We are now moving toward more interesting problems. What is the minimum number of layers you need to solve the two problems in Fig. 3? Your solution should contain combinations of simple perceptrons (some of them computing logical functions).

2. Consider the two neuron linear network

    $$\tau \frac{dx_1}{dt} + x_1 = b_1 + W_{11}x_1 + W_{12}x_2 \tag{2}$$

    $$\tau \frac{dx_2}{dt} + x_2 = b_2 + W_{21}x_1 + W_{22}x_2 \tag{3}$$

    with

    $$W = \begin{pmatrix} -1/4 & 1 \\ 1/2 & 1/4 \end{pmatrix} \tag{4}$$

    (a) Compute the steady state response to $b_1 = 1$ and $b_2 = 2$ by solving the linear equation $(I - W)x = b$ for x.

    (b) Find the eigenvalues of W. What are the gains of the corresponding eigenmodes?

    (c) Find the left and right eigenvectors of W. Because W is not symmetric, the left and right eigenvectors are different from each other. You'll see common and differential modes, but they won't be as simple as what we derived in class for the mutually inhibitory pair.

    (d) Use the eigenvectors and eigenvalues to factorize W as $S\Lambda S^{-1}$.

(e) Now compute the steady state response in the following three steps. You should get the same answer as you got from solving $(I - W)x = b$. This is because $(I - W)^{-1} = S(I - \Lambda)^{-1}S^{-1}$.

    i. Express b as a linear combination of right eigenvectors.

    ii. Multiply the coefficients of this linear combination by the appropriate gains.

    iii. Use these results to recombine the right eigenvectors, yielding the steady state response x.

3. Lateral inhibition and the horseshoe crab retina

A simple model of the horshoe crab retina is a one-dimensional chain of neurons (considered infinite here to avoid border effects) with inhibition between nearest neighbors (see Fig. 4a). It is governed by the following equation:

$$\frac{dx_i}{dt} + x_i = b_i - \beta x_{i+1} - \beta x_{i-1} \tag{5}$$

A more realistic model would have two dimensions, and longer range inhibition. However, the simple model goes a long way to explaining the phenomenon of edge enhancement.

(a) What are the eigenmodes of this network, and the corresponding eigenvalues/gains? What condition is necessary for the network to be stable?

Hint: Define the Nth root of unity $z = exp(2\pi i/N)$ and the Fourier matrix $F_{mn} = z^{mn}$. the Fourier modes are defined as column vectors of F. Show that the Fourier modes are eigenvectors of W and find the eigenvalues of these modes.

(b) In general, it is difficult to compute the neurons' response. However for small β values, it is possible to 'linearize' the system. Figure 4b depicts such an ideal result after stimulating the model with an input pattern that is a step from b_L to b_R.

    i. Away from the edge:
    In the following, we assume that all neurons away from the edge have the same response profile. Taylor expand the matrix W and show that the response of the neurons is a step from $x_L = (1-2\beta)b_L$ to $x_R = (1 - 2\beta)b_R$.
    Hint: $(1 + X)^{-1} \approx 1 - X$.

    ii. At the edge:
    Show that the response of the neuron just left of the edge is enhanced to

$$x_L + \beta(b_L - b_R)$$

    and the response of the neuron just right of the edge is similarly suppressed.

(c) Simulate the model. Plot the output profile at steady state and submit your plot. Explain what is happening at the borders of your network.

Use the following network parameters: number of neurons $= 20$, $b_L = 0.7$ and $b_R = 0.3$. Submit two plots corresponding to two β values, one for which the approximation is good and one for which it breaks down. This is because of disinhibition...

(d) Disinhibition

Two wrongs don't make a right, but the effect of two inhibitions can resemble excitation. This phenomenon is called disinhibition. Here is an illustration. Assume neuron C inhibits B, but is far enough from A that it has no direct inhibitory effect. Suppose we stimulate A and B. In the middle of the recording, C is stimulated also. This causes the firing rate of B to decrease, and the firing rate of A to increase. This presumably occurs because the inhibitory effect of B on A has been reduced.

On the other hand, stimulating C alone has no effect on A if B is not also being stimulated. This is because B cannot mediate the disinhibitory effect when it is inactive. This is a nonlinear effect, due to the threshold for firing action potentials.

Disinhibition can also been seen in the model with nearest neighbor inhibition. Suppose that all the elements of b are zero, except for $b_0 = 1$.   2

i. Show that the steady state response is

$$x_i = \frac{\gamma^{|i|}}{(1 + 2\beta)\gamma}$$

where

$$\gamma = \frac{\sqrt{1 - 4\beta^2} - 1}{2\beta}$$

If β is positive, then γ is negative, which means that x_i alternates between positive and negative values. The positive values are the result of disinhibition, disdisdisinhibition, and so on.

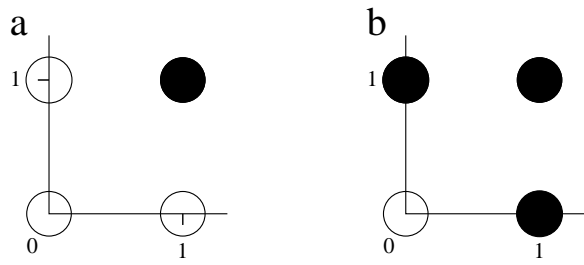ii. Simulate the network and submit a plot that contains both empirical and theoretical output profiles.
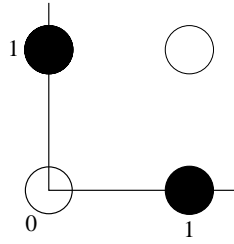
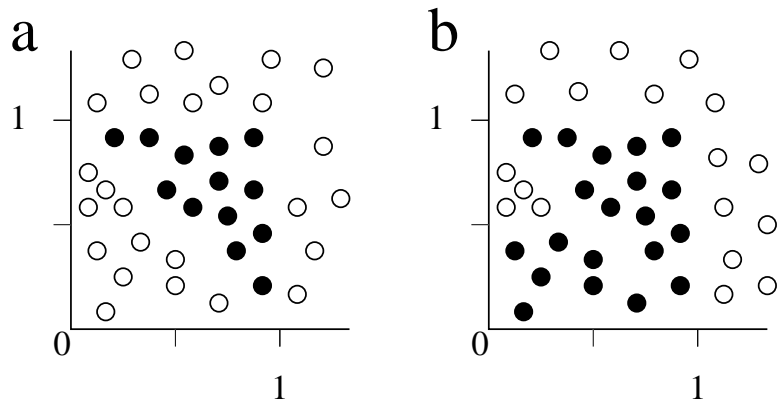Figure 1: Simple binary cases

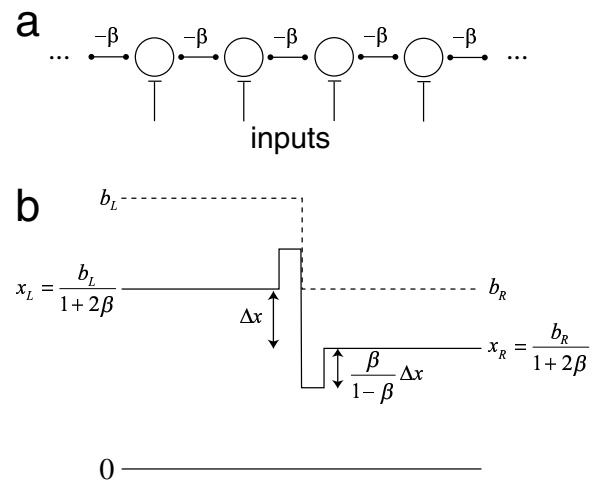
Figure 2: XOR

Figure 3: More complex problems

4

Figure 4: One-dimensional model with nearest neighbor inhibition. (a) Architecture of model. (b) The response to an edge shows enhancement.