E9 253: Neural Networks and Learning Systems - I **Fall 2021**

Homework #1

Course Instructor: Prof. Shayan Srinivasa Garani *Scribes: Kamal Bhatt, Sujeet Kumar Maurya*

Question 1:(a) How would you separate all integers modulo N on a real line using linear decision boundaries? (b) Explain in your own words (no more than 50 words) the stochastic resonance effect in neurons.

Solution(a): The set of integers can be partitioned based on the modulo N congruence, for example:

$$[0] = \{0, N, 2N, \dots\} = \{N\mathbb{Z}\}.$$

$$[1] = \{1, N+1, 2N+1, \dots\} = \{N\mathbb{Z}+1\}.$$

$$[2] = \{2, N+2, 2N+2, \dots\} = \{N\mathbb{Z}+2\}.$$

$$[3] = \{3, N+3, 2N+3, \dots\} = \{N\mathbb{Z}+3\}.$$

.

.

$$[N-1] = \{N-1, 2N-1, \dots\} = \{N\mathbb{Z}+(N-1)\}.$$

So, we have a natural way to classify the set of integers modulo N into N classes.
Consider any $a \in \mathbb{Z}$ then 'a' can be expressed as:

$$a \equiv b \bmod N, \text{ where } b \in \{0, 1, 2, \dots, N-1\}.$$

So, we now consider the transformation of f: $\mathbb{R} \to \mathbb{R}^2$

$$f(a) \mapsto (a, b)$$

Now we can define our linear decision boundaries. The linear decision boundaries are the set

$$\{y = i - 0.5 | 1 \le i \le N-1\}.$$

Solution(b): Stochastic resonance effect: In most cases, random noise leads to a drop in the quality of signal transmission, detection and performance. However, at times an increase in unpredictable fluctuations might instead increase the signal-to-noise ratio (SNR), particularly in neurons. This counter-intuitive effect is known as stochastic resonance effect. In other words, it is a phenomenon in which a signal that is normally too weak to be detected can have its detection drastically improved by adding noise containing a broad frequency range. Those noise frequencies that match with corresponding frequencies in the original signal resonate with the neural system, amplifying the original signal.

Question 2: Sketch the architecture of a single hidden layer recurrent network with 2 input nodes, 2 hidden nodes and an output node. Self loops and lateral connections are not allowed. Assuming the stochastic neuron model, write down the equations for the signals at the output of each neuron. Indicate all the necessary variables carefully.

Solution: In a recurrent neural network, the outputs are used as a feedback to the neurons of the previous layer. So,

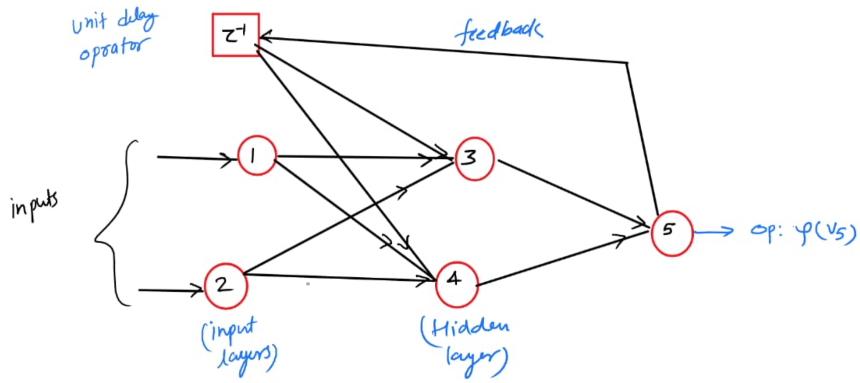

Figure 2.1: Network architecture.

It is to be seen that the above recurrent neural network has no self-loops or lateral connections.
Let $v_1(t), v_2(t), v_3(t), v_4(t)$ and $v_5(t)$ be outputs of each neuron respectively. Similarly, let $u_1(t), u_2(t), u_3(t), u_4(t)$ and $u_5(t)$ be the inputs of each neuron. We have

$$v_i(t) = \varphi(u_i(t)) \tag{2.1}$$

Let $x_1(t), x_2(t), x_3(t), x_4(t)$ and $x_5(t)$ be the state/firing of the neurons respectively given by

$$x_i(t) = \begin{cases} 1, & \text{with probability} = p_i(t) \\ 0, & \text{with probability} = 1 - p_i(t) \end{cases}$$

where $p_i(t) = f(u_i, t) = \left(\frac{1}{1 + e^{u_i/t}}\right)$
u_1 and u_2 are two inputs.

$$u_3(t) = x_1(t)v_1(t) + x_2(t)v_2(t) + x_5(t-1)v_5(t-1) \tag{2.2}$$
$$u_4(t) = x_1(t)v_1(t) + x_2(t)v_2(t) + x_5(t-1)v_5(t-1) \tag{2.3}$$
$$u_5(t) = x_3(t)v_3(t) + x_4(t)v_4(t) \tag{2.4}$$

Equations (2.1), (2.2), (2.3) and (2.4) are the state equations.
The signals have been written without considering weights; if they are considered, they will be added as coefficients to the terms of the equations (2.1), (2.2), (2.3) and (2.4).

Question 3: The output response of a certain device obeys the law $y = \frac{x}{a + bx}$, where a, b are positive constants. You make measurements $(x_i,\ y_i)$ over N data points that could be potentially noisy. Nothing is known about the statistics of the noise. You are required to fit the parameters a and b empirically from the data. How would you accomplish this? Provide explicit expressions for your estimates of a and b using techniques learnt in the class.

Solution: The output response of a certain device obeys the law

$$y = \left(\frac{x}{a+bx}\right).$$

Let $Y = \left(\frac{1}{y}\right)$ and $X = \left(\frac{1}{x}\right)$.

From the above law,

$$\begin{aligned} \left(\frac{1}{y}\right) &= \left(\frac{a}{x}+b\right). \\ \implies \quad Y &= aX + b. \end{aligned}$$

Also, given that (x_i, y_i), $1 \le i \le N$ data points could be potentially noisy, we formulate

$$Y = aX + b + \mathcal{E},$$

where $\mathcal{E}$ is the sample noise. So, the residual square sum (RSS) is computed as

$$RSS = \sum_{i=1}^{N}(Y_i - (aX_i + b + \epsilon_i))^2,$$

where $$Y_i = \frac{1}{y_i} \quad \text{and } X_i = \frac{1}{x_i}.$$

Goal: Minimize RSS over a, b

So, taking the partial derivatives w.r.t a and b and setting them to zero, we have

$$\frac{\partial(RSS)}{\partial a} = 0 \quad \text{and} \quad \frac{\partial(RSS)}{\partial b} = 0.$$

$$\begin{aligned} \implies \quad \frac{\partial(RSS)}{\partial b} &= 0. \\ \implies \quad -2\sum_{i=1}^{N}(Y_i - (aX_i + b + \epsilon_i)) &= 0. \\ \implies \quad \sum_{i=1}^{N} Y_i &= Nb + a\sum_{i=1}^{N} X_i + a\sum_{i=1}^{N} \epsilon_i. \\ \implies \quad b &= \frac{1}{N}\sum_{i=1}^{N} Y_i - a\frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{N}\sum_{i=1}^{N} \epsilon_i. \end{aligned}$$

$$\boxed{b = \overline{Y} - a\overline{X} - \overline{\mathcal{E}}} \tag{2.5}$$

where, $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{y_i}$, $\quad \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i} \quad$ and $\quad \overline{\mathcal{E}} = \text{mean}(\mathcal{E})$.

Now,

$$\frac{\partial(RSS)}{\partial a} = -2\sum_{i=1}^{N}(Y_i - (aX_i + b + \epsilon_i))X_i = 0.$$

$$\implies \qquad \sum_{i=1}^{N} X_iY_i = a\sum_{i=1}^{N} {X_i}^2 + \sum_{i=1}^{N}\epsilon_iX_i + b\sum_{i=1}^{N} X_i.$$

Putting the value of b from equation (2.5),

$$\implies \qquad \sum_{i=1}^{N} X_iY_i = a\sum_{i=1}^{N} {X_i}^2 + (\overline{Y} - a\overline{X} - \overline{\mathcal{E}})\sum_{i=1}^{N} X_i + \sum_{i=1}^{N}\epsilon_iX_i.$$

$$\implies \qquad a(\sum_{i=1}^{N} {X_i}^2 - \overline{X}\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} X_iY_i - \overline{Y}\sum_{i=1}^{N} X_i + \overline{\mathcal{E}}\sum_{i=1}^{N} X_i - \sum_{i=1}^{N}\epsilon_iX_i.$$

$$\boxed{a = \frac{\sum_{i=1}^{N} X_iY_i - \overline{Y}\sum_{i=1}^{N} X_i + \overline{\mathcal{E}}\sum_{i=1}^{N} X_i - \sum_{i=1}^{N}\epsilon_iX_i}{\sum_{i=1}^{N} {X_i}^2 - \overline{X}\sum_{i=1}^{N} X_i}} \tag{2.6}$$

Now, numerator of equation (2.6) can also be written as

$$\sum_{i=1}^{N} X_iY_i - N\frac{1}{N}\overline{Y}\sum_{i=1}^{N} X_i + N\overline{\mathcal{E}}\frac{1}{N}\sum_{i=1}^{N} X_i - \sum_{i=1}^{N}\epsilon_iX_i.$$

$$\implies \quad \sum_{i=1}^{N}(Y_i - \epsilon_i)X_i - N\overline{X}\ \overline{Y} + N\overline{X}\ \overline{\mathcal{E}}.$$

$$\implies \quad \sum_{i=1}^{N} X_i(Y_i - \epsilon_i) - (\overline{Y} - \overline{\mathcal{E}})\sum_{i=1}^{N} X_i - \overline{X}\sum_{i=1}^{N}(Y_i - \epsilon_i) + \sum_{i=1}^{N}\overline{X}(\overline{Y} - \overline{\mathcal{E}})$$

$$\implies \quad \sum_{i=1}^{N}(X_i - \overline{X})((Y_i - \epsilon_i) - (\overline{Y} - \overline{\mathcal{E}})).$$

$$\implies \quad \sum_{i=1}^{N}(X_i - \overline{X})((Y_i - \overline{Y}) - (\epsilon_i - \overline{\mathcal{E}})).$$

Similarly, denominator of equation (2.6) can be written as

$$\sum_{i=1}^{N} {X_i}^2 - \overline{X}\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} {X_i}^2 - N\overline{X}^2 = \sum_{i=1}^{N}(X_i - \overline{X})^2.$$

Hence, equation (2.6) can be written as

$$\boxed{a = \frac{\sum_{i=1}^{N}(X_i - \overline{X})((Y_i - \overline{Y}) - (\epsilon_i - \overline{\mathcal{E}}))}{\sum_{i=1}^{N}(X_i - \overline{X})^2}} \tag{2.7}$$

Putting the value of a in equation (2.5)

$$\boxed{b = \overline{Y} - \frac{\sum_{i=1}^{N}(X_i - \overline{X})((Y_i - \overline{Y}) - (\epsilon_i - \overline{\mathcal{E}}))}{\sum_{i=1}^{N}(X_i - \overline{X})^2}\overline{X} - \overline{\mathcal{E}}} \tag{2.8}$$

Note: Above fitting of data is valid assuming $x_i, y_i \neq 0 \ \ \forall$ i.

Question 4: Consider a 3-class classification problem, comprising labels $w_i, i = 1, 2, 3$ corresponding to data points distributed as shown in Figure 1 supported over [-3, 0], [-1.5, 1.5] and [0, 3] respectively. The corresponding *apriori* probabilities for the classes are $\frac{1}{2}$,$\frac{1}{3}$ and $\frac{1}{6}$. Are the points linearly separable? Determine the optimum thresholds and provide a Bayes decision rule to decide the label for a point randomly sampled from the interval [-3, 3]. Compute the probability of misclassification error.

Solution: The *apriori* probabilities for classes w_1, w_2, w_3 are
P(w_1)=$\frac{1}{2}$,P(w_2)=$\frac{1}{3}$,P(w_3)=$\frac{1}{6}$.
The general form of triangular distribution over the support [a,c] is given by

$$\mathbb{P}(x) = \begin{cases} 0, (x < a) \\ \dfrac{2(x-a)}{(b-a)(c-a)}, (a \leq x < c) \\ \dfrac{2(b-x)}{(b-a)(b-c)}, (c \leq x \leq b) \end{cases}$$

Hence, the conditional likelihoods given the classes w_1, w_2, w_3 is computed as,

$$\mathbb{P}(x \mid w_1) = \begin{cases} \dfrac{4(x+3)}{9}, x \in \left(-3, -\frac{3}{2}\right) \\ \left(-\dfrac{4x}{9}\right), x \in \left(-\frac{3}{2}, 0\right) \end{cases}$$

$$\mathbb{P}(x \mid w_2) = \begin{cases} \dfrac{2(2x+3)}{9}, \ x \in \left(-\frac{3}{2}, 0\right) \\ \dfrac{2(3-2x)}{9}, \ x \in \left(0, \frac{3}{2}\right) \end{cases}$$

$$\mathbb{P}(x \mid w_3) = \begin{cases} \left(\dfrac{4x}{9}\right), \ x \in \left(0, \frac{3}{2}\right) \\ \dfrac{4(3-x)}{9}, \ x \in \left(\frac{3}{2}, 3\right) \end{cases}$$

For threshold calculations,
1. In $\left[-\dfrac{3}{2}, 0\right]$, for overlapping region for w_1 and w_2;

$$P(x|w_1).P(w_1) = P(x|w_2).P(w_2).$$
$$\implies -\left(\frac{4x}{9}\right)\left(\frac{1}{2}\right) = \left(\frac{2(2x+3)}{9}\right)\left(\frac{1}{3}\right).$$
$$\implies x = -\left(\frac{3}{5}\right).$$

2. Similarly in $\left[0, \frac{3}{2}\right]$, for overlapping region for w_2 and w_3,

$$P(x|w_2).P(w_2) = P(x|w_3).P(w_3).$$
$$\implies \left(\frac{2(3-2x)}{9}\right)\left(\frac{1}{3}\right) = \left(\frac{4x}{9}\right)\left(\frac{1}{6}\right).$$
$$\implies x = 1.$$

So, the decision boundaries for classification would be

$$\text{for } x \in \left[-3,\ -\frac{3}{5}\right] \rightarrow w_1$$
$$\text{for } x \in \left[-\frac{3}{5},\ 1\right] \rightarrow w_2$$
$$\text{for } x \in [1,\ 3] \rightarrow w_3$$

Probability of misclassification error:

$$P(\text{error}) = \int_{-\frac{3}{2}}^{-\frac{3}{5}} P(x|w_2).P(w_2)\,dx + \int_{-\frac{3}{5}}^{0} P(x|w_1).P(w_1)\,dx + \int_{0}^{1} P(x|w_3).P(w_3)\,dx + \int_{1}^{3} P(x|w_2).P(w_2)\,dx.$$

By substituting the values, we have

$$P(\text{error}) = \int_{-\frac{3}{2}}^{-\frac{3}{5}} \frac{2(3-2x)}{9}\frac{1}{3}\,dx + \int_{-\frac{3}{5}}^{0} \left(-\frac{4x}{9}\right)\frac{1}{2}\,dx + \int_{0}^{1} \frac{4x}{9}\left(\frac{1}{6}\right)\,dx + \int_{1}^{3} \frac{2(3-2x)}{9}\left(\frac{1}{3}\right)\,dx.$$

$$P(\text{error}) = 0.156$$