NNSP-1

Homework #1

Instructor: Prof., Shayan Garani Srinivasa *TA: Prayag Gowgi*

Problem 1

Understand supervised, unsupervised and reinforced learning mechanisms (refer to any material). Which learning mechanism would you suggest in the following scenarios?

(a) Imagine yourself working for Flipkart or Amazon and you would like to recommend a book to a user.

(b) You would like to develop an android based game, say chess.

(c) You are interested in learning to play guitar.

(d) You are developing an application for a bank that decides whether to give a loan or not to a customer.

Solution

(a) Supervised learning/Unsupervised.

(b) Reinforced learning.

(c) Supervised learning.

(d) Supervised learning.

Note : You should justify your answers.

Problem 2

We humans tend to forget events with time unless it is reinforced again and again. There are exceptional individuals who can remember almost every event. Let us ignore such cases for a moment. Memory can be modeled using a feedback loop within a neural network architecture. Forgetting events in humans can be modeled as a reduction in the feedback strength with time. Justify the statement mathematically for a single neuron case.
Hint : In the class, we learned signal flow graphs of a single-loop feedback system.

Solution

Consider a single neuron with feedback loop as shown in Figure 1. The Z- transform of the output $y(n)$ is given by,

$$Z(y(n)) = \frac{w}{1 - wz^{-1}} Z(x(n)). \tag{1}$$

Equation (1) can be written as

$$Z(y(n)) = w \sum_{l=0}^{\infty} \left(wz^{-1}\right)^{l} Z(x(n)), \tag{2}$$

using the properties of Z- transform in equation (2) we get

$$y(n) = \sum_{l=0}^{\infty} w^{l+1} \underbrace{x(n-l)}_{\text{past information}} . \tag{3}$$

From equation (3), we see that the term $x(n-l)$ for $l = 0, \ldots, \infty$ refers to the past signals (memory). However, forgetting in humans implies that as $l \to \infty$, the contribution of $x(n-l)$ should go to zero. As a consequence, we require $|w| < 1$.

Problem 3

In the class, we learned that biological neurons are slower compared to artificial neurons. However, we are able to do more complex tasks in lesser time compared to artificial neuron based systems. What do you think is the reason behind this? (Imagine that you are given the same number of biological and artificial neurons using Silicon gates.)

Solution

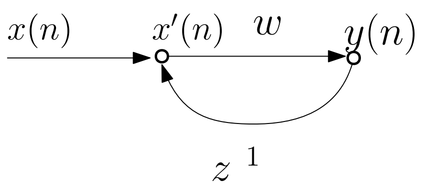


FIGURE 1 – Single neuron with feedback loop

(a) Highly power efficient.
(b) Massively interconnected.
(c) Parallel processing.
(d) stores information in the form of memory.
(e) Whenever an input is presented to the network of neurons, the input is given to all the neurons at the same instant. A particular neuron will respond strongly resulting in recall.

Problem 4

Consider the sigmoid activation function. The parameter a is called the slope parameter.

$$\phi(v) = \frac{1}{1 + \exp\left(-av + b\right)} \tag{4}$$

(a) With $b = 0$, plot the sigmoid activation function for different values of a (for $a = 0, 1, 10, 100, 200$).
(b) With $a = 1$, plot sigmoid activation function for different values of bias $b = -10, 0, 10$. What is the role of bias in the neuronal model ?

Solution

(a) The slope of sigmoid function increases with $a = 0, 1, 10, 100, 200$. Refer Figure 2(a).
(b) The sigmoid function is shifted along the v axis. Bias can be used to control the firing threshold of a neuron. Refer Figure 2(b).

(a) Sigmoid activation function for different slope parameter **a**.

(b) Sigmoid activation function for different bias values.

FIGURE 2 – Expected plots for Question 4

Problem 5

An odd sigmoid function is defined by

$$\phi(v) = \frac{1 - \exp(-av)}{1 + \exp(-av)} = \tanh\left(\frac{av}{2}\right), \tag{5}$$

where tanh(.) is the hyperbolic tangent function. Show that the derivative of $\phi(v)$ with respect to v is given by

$$\frac{d\phi}{dv} = \frac{a}{2}\left(1 - \phi^2(v)\right). \tag{6}$$

What is the value of this derivative at the origin ? Suppose that the slope parameter a is made ∞. What is the resulting form of $\phi(v)$?

Solution

Differentiating equation (5) with respect to v we get

$$\begin{aligned}\phi'(v) &= \frac{2a\exp(-av)}{(1+\exp(-av))^2} & (7)\\ &= \frac{2a\exp\left(-\frac{av}{2}\right)^2}{\exp\left(-\frac{av}{2}\right)^2\left(\exp\left(\frac{av}{2}\right)+\exp\left(-\frac{av}{2}\right)\right)^2} & (8)\\ &= \frac{2a\times 4}{4\left(\exp\left(\frac{av}{2}\right)+\exp\left(-\frac{av}{2}\right)\right)^2} & (9)\\ &= \frac{a}{2}\left(1-\phi^2(v)\right) & (10)\end{aligned}$$

At $v=0$, $\phi'(v)=0$ and as $a\to\infty$, $\phi(v)\to 1$.

Problem 6

Understand the cumulative distribution function (CDF) (refer to any material). Which of the following sigmoid functions qualifies as a cumulative distribution function ? Justify your answer.

$$\phi(v) = \frac{1}{1+\exp(-av)} \quad (11)$$

$$\phi(v) = \frac{1-\exp(-av)}{1+\exp(-av)} \quad (12)$$

$$\phi(v) = \frac{v}{\sqrt{1+v^2}} \quad (13)$$

$$\phi(v) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{v}\exp\left(-\frac{x^2}{2}\right)dx \quad (14)$$

$$\phi(v) = \frac{2}{\pi}\arctan(v) \quad (15)$$

Solution

Cumulative distribution function (CDF) of random variable X is the probability that the random variable X is less than or equal to some real number x i.e.,

$$F_X(x) = P(X \le x) \quad (16)$$

and satisfies the following properties :

(a) right continuous

(b) non-decreasing.

(c) $\lim_{x\to+\infty} F_X(x) = 1$.

(d) $\lim_{x\to-\infty} F_X(x) = 0$.

1. Equation (11) satisfies conditions (a)-(d). It is a valid CDF.
2. Equation (12) violates condition (d). It is not a valid CDF.
3. Equation (13) violates condition (d). It is not a valid CDF.
4. Equation (14) is valid CDF.
5. Equation (15) violates condition (d). It is not a valid CDF.

Problem 7

In the class, we learned the model of an artificial neuron as shown in Figure 3.

(a) The probability of a stochastic neuron firing is given by

$$P(v) = \frac{K}{1+\exp\left(-\frac{v}{T}\right)}, \quad (17)$$

where $T\in[0,\infty]$ is a control parameter, K is a constant (refer to your class notes). For what value of K, is $\phi(v)$ a valid probability density function (pdf) ?

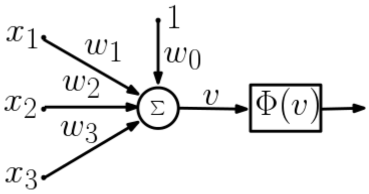


FIGURE 3 – Artificial neuron model

(b) Using the neuronal model in question 7 (refer Figure 3), come up with a feed forward neural network architecture with two neurons. (Use any drawing tool to draw the architecture). Write down the equations at each and every point in the network.

(c) Assuming the stochastic neuron model as given in question 7(a) in your feed forward architecture, write down the equations at each and every point within your architecture.

(d) With the stochastic neuron model as given in question 7(a), come up with a recurrent neural network architecture with no self feedback loops on each neuron. Write all the equations at each and every point in your architecture.

(e) In the class, we learned that the output of a stochastic neuron is binary i.e.,

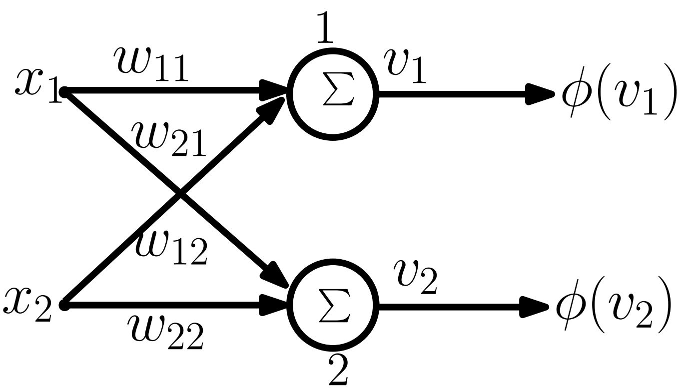
$$X = \begin{cases} +1, & \text{with probability } \; P(v) \\ -1, & \text{with probability } 1 - P(v). \end{cases} \tag{18}$$

If you require the output of a stochastic neuron to be continuous rather than binary, for some application. How would you modify your neuron model ?

Solution

(a) For no value of K, equation (17) is a valid pdf. However, by differentiating equation (17) and integrating it from $-\infty$ to ∞ would result in $K = +1$ for the equation (17) to be a valid pdf.

(b) The architecture is shown in Figure 4.



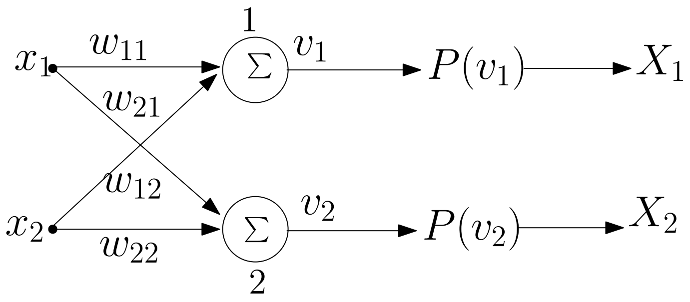
$$v_1 = x_1 w_{11} + x_2 w_{21}$$
$$v_2 = x_1 w_{12} + x_2 w_{12}$$

FIGURE 4 – Feed forward architecture with two neurons.

(c) The architecture is shown in Figure 5. From Figure 5 the outputs X_1 and X_2 are given by,

$$X_1 = \begin{cases} +1, & \text{with probability } \; P(v_1) \\ -1, & \text{with probability } 1 - P(v_1). \end{cases} \tag{19}$$

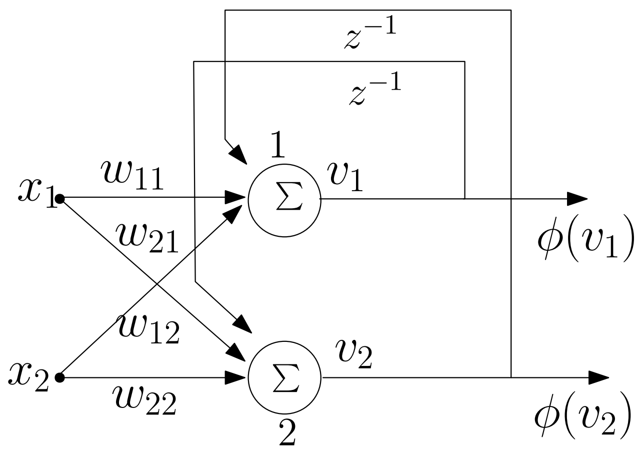
$$X_2 = \begin{cases} +1, & \text{with probability } \; P(v_2) \\ -1, & \text{with probability } 1 - P(v_2). \end{cases} \tag{20}$$



$$v_1 = x_1 w_{11} + x_2 w_{21}$$
$$v_2 = x_1 w_{12} + x_2 w_{12}$$

FIGURE 5 – Feed forward architecture with two stochastic neurons.

(d) The architecture is shown in Figure 6.



$$v_1(n) = x_1(n) w_{11}(n) + x_2(n) w_{21}(n) + v_1(n-1)$$
$$v_2(n) = x_1(n) w_{21}(n) + x_2(n) w_{12}(n) + v_2(n-1)$$

FIGURE 6 – Recurrent neural network architecture.

(e) Define an invertible continuous function at the output of stochastic neuron with $P(v)$ as its domain. For example, a simple linear function i.e.,

$$y = P(v) \tag{21}$$