INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Mid-Autumn Semester 2022-23

Duration: 2 hrs   Full Marks: 60

Subject No: ES60011   Subject: Application Of Machine Learning In Biological Systems
Department/Center/School: Energy Science and Engineering
Specific charts, graph paper, log book etc., required: None
Special Instructions (if any): *(1) Answer all the questions. (2) In case of reasonable doubt, make practical assumptions and write that on your answer script. (3) The parts of each question must answered be together. (4) Calculator is allowed.*

---

1. Justify the following statements stating whether it is true/false. Justification is must.

    (a) Cross validation is required to verify the correctness of the result.
    (b) Data balancing is one of the key to the success of any machine learning method.
    (c) Protein secondary structure can be predicted from the protein sequence only.
    (d) For protein phosphorylation site prediction problem, data balancing can be done considering some (which?) biological insight.

    Marks: 3+3+3+3=12

---

2. TP=10000 ; FP=400; FN=300; TN=1000

    A machine learning algorithm generates above mentioned result.
    (a) Compute the accuracy of the method.
    (b) Compare the accuracy with F1-score. F1 score [F=2×(P ×R)/(P+R)] is the harmonic mean of precision and recall.

    Marks: 3+5=8

---

3. Consider the following Multiple Sequence Alignment of proteins as output from Clustal Omega software. Compute the Henikoff Weight for each of the sequences

 | 7PUB_10 | | QKMLQRKVTCFQ |
    |---|---|---|
    | 12AS_1 | | DRLSPLHSVYVD |
    | 11AS_1 | | DRLSPLHSVYVD |
    | 6HIV_56 | | YFIVKRCTLYFS |

    Marks: 2.5×4=10

---

4. (a) Write down the steps of Principal Component Analysis in connection with feature analysis and dimension reduction.

(b) Compute the Eigen values for the following covariance matrix. Next compute the Eigen vector corresponding to the highest Eigen value.

$$\begin{pmatrix} 2 & 2 \\ 5 & -1 \end{pmatrix}$$

Marks: 4+(4+2)=10

5.

Input Layer   Hidden Layer   Output Layer
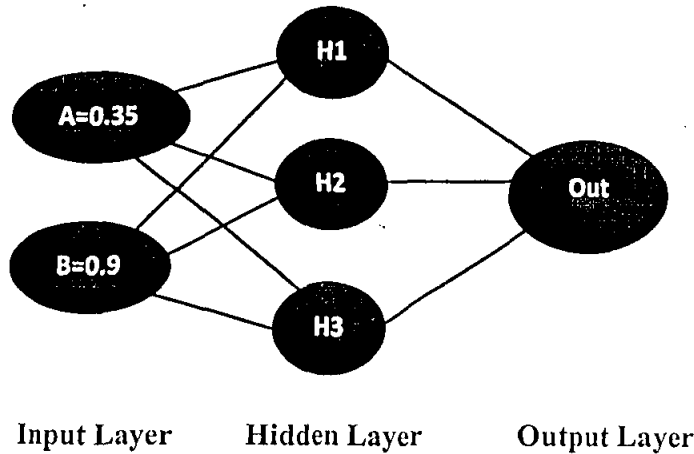
Initial edge weights:

AH1=0.3 ; AH2=0.2 ; AH3=0.4

BH1=0.3 ; BH2=0.4 ; BH3=0.2

H1Out=0.2 ; H2Out=0.3 ; H3Out=0.4

Assume that the neurons have a Sigmoid activation function

(a) Perform a forward pass on the network and compute the values at each hidden and output node.

(b) Perform a reverse pass (training) once (target=0.5) and compute the new edge weights by clearly marking your error.

(c) Perform a further forward pass with the new edge weights as computed in (b) to compute the modified output value. Comment on the result.

Please show steps of the calculations. Only final answer will not get any marks.

Marks: 5+10+5=20