IIT Kanpur Intensive Training School (ITS) on PYTHON for Machine Learning, Neural Networks and Deep Learning

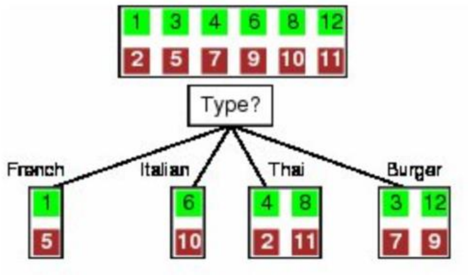
2nd to 22nd December 2023

Assignment #2

1. The centroid of the ith cluster in lth iteration, denoted by $\bar{\boldsymbol{\mu}}_i^{(l)}$, is
   a. $\frac{\sum_{j=1}^{M} \alpha_i^{(l)}(j)\bar{\mathbf{x}}(j)}{M}$
   b. $\frac{\sum_{j=1}^{M} \bar{\mathbf{x}}(j)}{M}$
   c. $\frac{\sum_{j=1}^{M} \bar{\mathbf{x}}(j)}{\sum_{j=1}^{M} \alpha_i^{(l)}(j)}$
   d. $\frac{\sum_{j=1}^{M} \alpha_i^{(l)}(j)\bar{\mathbf{x}}(j)}{\sum_{j=1}^{M} \alpha_i^{(l)}(j)}$
2. The **entropy** $H(X)$ of a source is defined as
   a. $-\sum_{i=1}^{n} p(x_i)\log_2\frac{1}{p(x_i)}$
   b. $\sum_{i=1}^{n} \frac{1}{p(x_i)}\log_2\frac{1}{p(x_i)}$
   c. $-\sum_{i=1}^{n} \frac{1}{p(x_i)}\log_2\frac{1}{p(x_i)}$
   d. $\sum_{i=1}^{n} p(x_i)\log_2\frac{1}{p(x_i)}$
3. Consider the table below showing ***joint probabilities*** of
   $X = \{\mathrm{IC}, \overline{\mathrm{IC}}\}$, $Y = \{\mathrm{CHOC}, \overline{\mathrm{CHOC}}\}$

	IC	$\overline{\mathbf{IC}}$
CHOC	$\frac{1}{6}$	$\frac{1}{3}$
$\overline{\mathbf{CHOC}}$	$\frac{1}{6}$	$\frac{1}{3}$

   The quantity $H(Y|X = \overline{IC})$ is given as
   a. 0.92
   b. 1
   c. 0.73
   d. 0.65
4. Consider the example



The quantity $H(Y|\text{Italian}) =$

a. 1
b. $\frac{1}{2}$
c. $\frac{1}{4}$
d. 0

5. The size of a typical neuron is
   a. **4 microns** to **100 picometres**
   b. **4 microns** to **100 nanometres**
   c. **4 microns** to **100 micrometres**
   d. **4 microns** to **100 femtometres**
6. There are approximately ______________ neurons in the human brains
   a. 10 billion
   b. 10 million
   c. 10 thousand
   d. 10 trillion
7. In comparison to silicon logic gates, ***Neurons*** ____________
   a. Have an equal speed of operation
   b. Depends on the type of Neuron
   c. Are Much faster
   d. Are Much slower
8. The activation function employed in the ***McCulloch–Pitts model*** of a neural network is
   a. Sigmoid function
   b. ReLU function
   c. Linear function
   d. Threshold function
9. Consider the loss function of a single neuron neural net given as

$$L(\bar{\mathbf{w}}_k) = \frac{1}{2}\left(y_k - \psi\left(\underbrace{\sum_{j=1}^{m} w_{kj}x_j + b_k}_{u_k}\right)\right)^2$$

The corresponding update rule for the weight vector is

a. $\bar{\mathbf{w}}_k(n+1) = \bar{\mathbf{w}}_k(n) - \eta_k\left(y - \psi\big(u_k(n)\big)\right)\psi'\big(u_k(n)\big)\bar{\mathbf{x}}$
b. $\bar{\mathbf{w}}_k(n+1) = \bar{\mathbf{w}}_k(n) + \eta_k\left(y - \psi\big(u_k(n)\big)\right)\bar{\mathbf{x}}$
c. $\bar{\mathbf{w}}_k(n+1) = \bar{\mathbf{w}}_k(n) + \eta_k\left(y - \psi\big(u_k(n)\big)\right)\psi'\big(u_k(n)\big)\bar{\mathbf{x}}$
d. $\bar{\mathbf{w}}_k(n+1) = \bar{\mathbf{w}}_k(n) + \eta_k\psi'\big(u_k(n)\big)\bar{\mathbf{x}}$

10. The ***ReLU function*** $f(x)$ is defined as
    a. x
    b. $\begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$
    c. $\frac{1}{1+e^{-x}}$
    d. $\begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$
11. In a two layer deep neural network, the gradient to update the weights of layer 1 during backpropagation is determined as
    a. $\frac{\partial}{\partial \mathbf{W}^{[1]}} L = \left(\left(\mathbf{W}^{[2]}\right)^T \phi^{[2]} \right) \times \left(\bar{\mathbf{x}}^{[0]}\right)^T$
    b. $\frac{\partial}{\partial \mathbf{W}^{[1]}} L = \left(\left(\mathbf{W}^{[2]}\right)^T \phi^{[2]} \odot \left(\psi^{[1]}\right)' \left(\bar{\mathbf{z}}^{[1]}\right) \right) \times \left(\bar{\mathbf{x}}^{[0]}\right)^T$
    c. $\frac{\partial}{\partial \mathbf{W}^{[1]}} L = \left(\left(\mathbf{W}^{[2]}\right)^T \phi^{[2]} \odot \left(\psi^{[1]}\right)' \left(\bar{\mathbf{z}}^{[1]}\right) \right) \times \bar{\mathbf{x}}^{[0]}$
    d. $\frac{\partial}{\partial \mathbf{W}^{[1]}} L = \left(\left(\mathbf{W}^{[2]}\right)^T \odot \left(\psi^{[1]}\right)' \left(\bar{\mathbf{z}}^{[1]}\right) \right) \times \left(\bar{\mathbf{x}}^{[0]}\right)^T$
12. In a K layer neural network, the size of the matrix $\mathbf{W}^{[l]}$ for one of the inner layers is $m \times n$, where m, n are respectively
    a. Number of neurons in layers $l, l-1$
    b. Number of neurons in layers $l-1, l$
    c. Number of neurons in layers l and number of inputs in layer 0
    d. Number of neurons in layers l and number of outputs in layer K
13. Convolutional neural nets are primarily suited for
    a. Gaussian datasets
    b. Purchase datasets
    c. Random datasets
    d. Images/ video datasets
14. Consider the simple image below

2	2	3	1
4	3	4	2
2	2	1	1
3	5	2	3

Max pooling with 2×2 filters and stride 2 leads to

a.

3	4
5	3

b.

4	2
5	3

c.

4	4
5	3

d.

4	4

4	4

15. Consider the simple image below

2	2	3	1
4	3	4	2
2	2	1	1
3	5	2	3

Average pooling with 2×2 filters and stride 2 leads to

a.

2.75	2.25
3	1.75

b.

2.75	2.5
3	1.75

c.

2.75	2.5
3.25	1.75

d.

2.75	2.5
3	1.5