1) Neural Network

$\therefore$ For final layer $\boxed{h(a) = a}$

$\therefore E_n(w) = \frac{1}{2}(y(x_n, w) - t_n)^2$

$y(x_n, w) = h(a_1^{[3]}) = a_1^{[3]}$.

- Calculate $\dfrac{\partial E_n(w)}{\partial a_1^{[3]}}$

$a_1^{[3]}$ is activation of output node $\qquad \dfrac{\partial E_n(w)}{\partial a_1^{[3]}} \equiv \delta_1^{[3]}$.

$$\frac{\partial E_n(w)}{\partial a_1^{[3]}} = \frac{1}{2}(a_1^{[3]} - t_n)^2$$

$$= \frac{\cancel{2}}{\cancel{2}}[a_1^{[3]} - t_n] \times (1)$$

$$\boxed{\frac{\partial E_n(w)}{\partial a_1^{[3]}} \equiv \delta_1^{[3]} = [a_1^{[3]} - t_n]} \text{ --- } 1$$

- Calculate $\dfrac{\partial E_n(w)}{\partial w_{12}^{[3]}}$

$\therefore$ Using the chain rule.

$$\boxed{\frac{\partial E_n(w)}{\partial w_{12}^{[3]}} = \frac{\partial E_n(w)}{\partial a_1^{[3]}} \times \frac{\partial a_1^{[3]}}{\partial w_{12}^{[3]}}} \text{ --- } 2$$

$$\therefore a_1^{[3]} = w_{11}^{[3]} z_1^{[2]} + w_{12}^{[3]} z_2^{[2]} + w_{13}^{[3]} z_3^{[2]}$$

$$\boxed{\frac{\partial a_1^{[3]}}{\partial w_{12}^{[3]}} = z_2^{[2]}} \text{ --- } 3$$

Sub 1 and 3 in 2

$$\boxed{\frac{\partial E_n(w)}{\partial w_{12}^{[3]}} = [a_1^{[3]} - t_n] z_2^{[2]}}$$

- Consider penultimate layers of node.
- Expression for $\frac{\partial E_n(w)}{\partial a_1^{[2]}}$. Use $\delta_1^{[3]}$.

$$\delta_1^{[3]} \equiv \frac{\partial E_n(w)}{\partial a_1^{[3]}}$$

Chain rule:-

$$\frac{\partial E_n(w)}{\partial a_1^{[2]}} = \underbrace{\boxed{\frac{\partial E_n(w)}{\partial a_1^{[3]}}}}_{\delta_1^{[3]}} * \frac{\partial a_1^{[3]}}{\partial a_1^{[2]}}$$

$$\therefore a_1^{(3)} = w_{11}^{(3)} h(a_1^{[2]}) + w_{12}^{[3]} h(a_2^{[2]}) + w_{13}^{[3]} h(a_3^{[2]})$$

$$\frac{\partial a_1^{[3]}}{\partial a_1^{[2]}} = \boxed{w_{11}^{[3]} h'(a_1^{[2]})}$$

$$\boxed{\therefore \frac{\partial E_n(w)}{\partial a_1^{[2]}} = \delta_1^{[3]} w_{11}^{[3]} h'(a_1^{[2]})} \quad \text{— } 1$$

- Use this to calc $\frac{\partial E_n(w)}{\partial w_{11}^{[2]}}$

$$\boxed{\frac{\partial E_n(w)}{\partial w_{11}^{[2]}} = \frac{\partial E_n(w)}{\partial a_1^{[2]}} \times \frac{\partial a_1^{[2]}}{\partial w_{11}^{[2]}}} \quad \text{— } 2$$

$$\therefore a_1^{[2]} = w_{11}^{[2]} h(a_1^{[1]}) + w_{12}^{[2]} h(a_2^{[1]}) + w_{13}^{[2]} h(a_3^{[1]})$$

$$\boxed{\frac{\partial a_1^{[2]}}{\partial w_{11}^{[2]}} = h(a_1^{[1]})} \quad \text{— } 3$$

1 and 3 in 2

$$\boxed{\frac{\partial E_n(w)}{\partial w_{11}^{[2]}} = \delta_1^{[3]} w_{11}^{[3]} h'(a_1^{[2]}) h(a_1^{[1]})}$$

X———X

Consider the weights from input layer.

- Expression for $\frac{\partial E_n(w)}{\partial a_1^{[1]}}$. Use set of $\delta_k^{(2)}$.

$$\frac{\partial E_n(w)}{\partial a_1^{[1]}} = \sum_{k=1}^{3} \frac{\partial E_n(w)}{\partial a_k^{[2]}} \times \frac{\partial a_k^{[2]}}{\partial a_1^{[1]}}.$$

$$\sum_{k=1}^{3} \frac{\partial E_n(w)}{\partial a_k^{[2]}} = \sum_{k=1}^{3} \delta_k^{[2]}.$$

$$a_k^{[2]} = \sum_{k=1}^{3} \sum_{i=1}^{3} w_{ki}^{[2]} z_i^{[1]}$$

$$\frac{\partial a_k^{[2]}}{\partial a_1^{[1]}} = h'(a_1^{[1]}) \sum_{k=1}^{3} w_{k1}^{[2]}$$

$$\boxed{\frac{\partial E_n(w)}{\partial a_1^{[1]}} = h'(a_1^{[1]}) \sum_{k=1}^{3} w_{k1}^{[2]} \delta_k^{[2]}} \quad (1)$$

- Use this to calculate $\frac{\partial E_n(w)}{\partial w_{11}^{[1]}}$

$$\frac{\partial E_n(w)}{\partial w_{11}^{[1]}} = \frac{\partial E_n(w)}{\partial a_1^{[1]}} \times \frac{\partial a_1^{[1]}}{\partial w_{11}^{[1]}}. \quad (2)$$

$$\therefore a_1^{[1]} = w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 + w_{13}^{[1]} x_3$$

$$\boxed{\frac{\partial a_1^{[1]}}{\partial w_{11}^{[1]}} = x_1} \quad (3)$$

Sub (1) & (3) in (2)

$$\boxed{\frac{\partial E_n(w)}{\partial w_{11}^{[1]}} = h'(a_1^{[1]}) \left[\sum_{k=1}^{3} w_{k1}^{[2]} \delta_k^{[2]}\right] x_1}$$

Fine-Tuning a Pre-Trained Network

Ran it for 2 Epochs as it was working very slow on my laptop.

Epoch 1 of 2
Training Set Validation
Loss: 1.1166619062423706
Accuracy: 0.6142797668609492
-------------------------------
Test Set Validation
Loss: 1.1166619062423706
Accuracy: 0.6142797668609492
-------------------------------
Epoch 2 of 2
Training Set Validation
Loss: 1.1194742918014526
Accuracy: 0.6060574521232306
-------------------------------
Test Set Validation
Loss: 1.1194742918014526
Accuracy: 0.6060574521232306
-------------------------------
Finished Training
##########################
----------Results---------
Loss in Training: 1.1166619062423706
Loss in Testing : 1.1166619062423706
Best Epoch from the total 2 Epochs: 1
##########################

Minimum Loss: 1.1166619062423706

Corresponding Epoch: 1

2. Bayesian Networks.

i a). $A \perp\!\!\!\perp C \mid B, D$.

When conditioned on B and D all the paths are blocked between A and C, hence this is $\boxed{\text{True}}$.

Mathematically.

$$\frac{P(A,B,C,D)}{P(B,D)} = \frac{P(A|B,D)\,P(\cancel{B|A,C})\,P(C|B,D)\,P(\cancel{D|A,C})}{P(\cancel{B|A,C})\,P(\cancel{D|A,C})}$$

$$= P(A|B,D)\,P(C|B,D)$$

Therefore $\underline{A \perp\!\!\!\perp C \mid BD \text{ is True}}$

---

. $B \perp\!\!\!\perp D \mid A, C$

When conditioned on A and C all the paths are blocked between B and D, hence this is $\boxed{\text{True}}$.

Mathematically.

$$\frac{P(A,B,C,D)}{P(A,C)} = \frac{\cancel{P(A|B,D)}\,P(B|A,C)\,\cancel{P(C|B,D)}\,P(D|A,C)}{\cancel{P(A|B,D)}\,\cancel{P(C|B,D)}}$$

$$= P(B|A,C)\,P(D|A,C).$$

$\therefore$ $\underline{B \perp\!\!\!\perp D \mid A,C \text{ is True}}$

---

Q2) Bayesian Networks.

ib) • $A \perp\!\!\!\perp C \mid B, D$.

This is $\boxed{\text{True}}$ because, when conditioned on B and D all paths b/w A and C are blocked. Arrows meet head to tail at node B and D.

$\therefore$ $A \perp\!\!\!\perp C \mid B, D$ holds True

• $B \perp\!\!\!\perp D \mid A, C$

This is $\boxed{\text{False}}$, when conditioned on ~~B and D~~ A and C there is still an active path b/w B and D.

Mathematically

$$\frac{P(A,B,C,D)}{P(A,C)} = \frac{\cancel{P(A)}\, P(B|A)\, P(D|A)\, P(\cancel{C|B,D})}{\cancel{P(A)} \times P(\cancel{C|B,D})}$$

$$= P(B|A)\, P(D|A)$$

$$\neq B \perp\!\!\!\perp D \mid A, C$$

$\therefore$ This is $\boxed{\text{False}}$

Q2 Bayesian Networks.

ic) • $A \perp\!\!\!\perp C \mid B, D$.

This is $\boxed{\text{False}}$ because even when B, D are observed there is still an active path b/w A and C as the path becomes unblock. In this case B and D are head to head.

Mathematically.

$$\frac{P(A,B,C,D)}{P(B,D)} = \frac{P(A)P(C)P(D|A,C)\cancel{P(B|A,C)}}{\cancel{P(B|A,C)}P(\cancel{D|A,C})}$$

$$\neq P(A)P(C)$$

$\therefore \neq A \perp\!\!\!\perp C \mid B, D$ is False

• $B \perp\!\!\!\perp D \mid A, C$

When conditioned on A & C all the path b/w B and D remain blocked and are therefore inactive.

Hence, this statement is $\boxed{\text{True}}$

Mathematically.

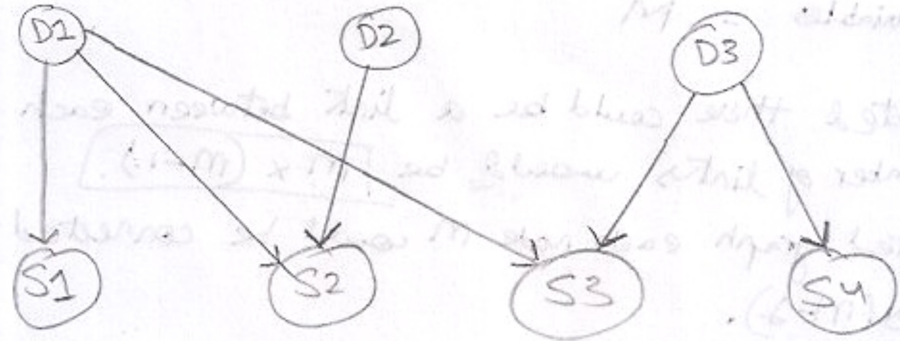
$$P(A,B,C,D) = \frac{\cancel{P(A)}\cancel{P(C)}P(B|A,C)P(D|A,C)}{\cancel{P(A)}\cancel{P(C)}}$$

$$= P(B|A,C)P(D|A,C)$$

$\therefore \boxed{B \perp\!\!\!\perp D \mid A, C \text{ is True}}$

2 Bayesian Networks.

i)a)




) $P(D1, D2, D3, S1, S2, S3, S4) =$

$$P(D1)P(D2)P(D3)\,P(S1|D1)\,P(S2|D1,D2)\,P(S3|D1,D3)\,P(S4|D3)$$

c) $P(D1)$ : 1     $P(S1|D1)$ : 2

$P(D2)$ : 1     $P(S2|D1)$ : 2

$P(D3)$ : 1     $P(S2|D2)$ : 2

$P(S3|D1)$ : 2

$P(S3|D3)$ : 2

$P(S4|D3)$ : 2

Total number of independent parameters.

$$1+1+1+2+2+2+2+2+2 = \boxed{15}$$

d) Total No. of Random Variable = 7

Assuming each parent is boolean that means they can take 2 values.

$2^K - 1$ [K = random variable]

$2^7 - 1 = 128 - 1 = \boxed{127}$ independent parameters would be required

3) Markov Random Fields.

Energy Function.

a) • $x_K$ is a target variable with two states $\{-1, 1\}$

• all the other variables are fixed.

$$E(x,y) = h\sum_{i\neq k} x_i - \beta\sum_{\{i,j\}\neq k} x_i x_j - \eta\sum_{i\neq k} x_i y_i + h x_K - \beta\sum_{l} x_K x_l - \eta x_K y_K$$

$$= h\sum_{i\neq k} x_i + h x_K - \beta\sum_{\{i,j\}\neq k} x_i x_j - \beta\sum_{l} x_K x_l - \eta\sum_{i\neq k} x_i y_i - \eta x_K y_K$$

• just like in $-\beta\sum_{\{i,j\}} x_i x_j$ $x_i$ and $x_j$ are adjacent and neighbours

Similarly $\boxed{-\beta\sum_{l} x_K x_l} \rightarrow$ here $x_K$ and $x_l$ are neighbours.

$\boxed{+h x_K - \beta\sum_{l} x_K x_l - \eta x_K y_K} \rightarrow$ shows dependence of $E(x,y)$ on $x_K$

$\therefore$ $E(x,y)$ can be written as.

$$\boxed{E(x,y) = h x_K - \beta\sum_{l} x_K x_l - \eta x_K y_K}$$

$x_K$ has two states $\{-1, 1\}$

To get change in energy sub $x_K = 1$ & $x_K = -1$ in $E(x,y)$ and get the difference

let $x_K = 1$

$E(x,y) = h - \beta\sum_{l} x_l - \eta y_K$.

let $x_K = -1$

$E(x,y) = -h + \beta\sum_{l} x_l + \eta y_K$

$$E(x,y) = \left(h - \beta\sum_{l} x_l - \eta y_K\right) - \left(-h + \beta\sum_{l} x_l + \eta y_K\right)$$

$$\boxed{E(x,y) = 2h - 2\beta\sum_{l} x_l - 2\eta y_K}$$

Hence this shows that the difference only depends on quantities that are local to $x_K$, which is implied by $h$, $x_l$ and $y_K$.

3) Markov Random Fields

b) $\beta = h = 0$

$$E(x,y) = 0 \sum_i x_i - 0 \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

$$\therefore E(x,y) = -\eta \sum_i x_i y_i$$

Most probable configuration is given by $x_i = y_i$ for all $i$

Suppose index $j$ satisfies $x_j \neq y_j$ which will result in

$x_j y_j = -1$.

By changing the sign of $x_i$ $E(x,y)$ could be minimized.

$\therefore$ For $y_i \in (-1,1)$ and $x_i \in (-1,1)$.

where $i = 1 \dots D$

To get the most probable configuration which maintains minimum energy in the energy function

set $\boxed{x_i = y_i \text{ for all } i}$

X——X

## Undirected Graph.

Total nodes = Random variables = M

As the graph is undirected there could be a link between each node pair. Total number of links would be $\boxed{m \times (m-1)}$.
As this is an undirected graph each node m could be connected to any of the other nodes $(m-1)$.
We need distinct links, therefore consider Node X and Node Y. As it is undirected there is a link from X to Y and Y to X which is being counted twice. TO get distinct links we would do $\boxed{\frac{m \times (m-1)}{2}}$

Number of distinct undirected graph would be 2 raised to the power of distinct links. that is:

$$\boxed{2^{\frac{m \times (m-1)}{2}}}$$

8 possibilities for $m = 3$ are :-