# CSCI3230 Fundamentals of Artificial Intelligence
## Written Assignment 3
### Suggested Solution

1. a) When I(V) = 0, V is in fact a constant as only one possibility is allowed,
When $I(V) = \log_2 n$, every possibility has an equal weight to occur, V contains the maximum information [5+5 marks]

b) Convert from Horn Clause to Conjunctive Normal Form

R1: $\neg American(x) \vee \neg Weapon(y) \vee \neg Sells(x, y, z) \vee \neg Hostile(z) \vee Criminal(x)$.

R2: $\neg Missile(x) \vee \neg Owns(Nono, x) \vee Sells(West, x, Nono)$.

R3: $\neg Enemy(x, America) \vee Hostile(x)$.

R4: $\neg Missile(x) \vee Weapon(x)$.

F5: $Owns(Nono, M_i)$.

F6: $Missile(M_i)$.

F7: $American(West)$.

F8: $Enemy(Nono, America)$. [10 marks]

Prove that West is a criminal: (Correct answer in other forms are also accepted)

Assume West is not a criminal: ~Criminal (West)

| Unification Input 1 | Unification Input 2 | Substitution | Result |
|---|---|---|---|
| $\sim American(West) \vee \sim Weapon(y) \vee \sim Sells(West, y, z) \vee \sim Hostile(z) \vee Criminal(West)$ | $\sim Criminal(West)$ | x=West | $\sim American(West \vee \sim Weapon(y) \vee \sim Sells(West, y,$ $\vee \sim Hostile(z)$ |
| $American(West)$ | $\sim American(West) \vee \sim Weapon(y) \vee \sim Sells(West, y, z) \vee \sim Hostile(z)$ | | $\sim Weapon(y) \vee \sim Sells(West, y,$ $\vee \sim Hostile(z)$ |
| $Enemy(Nono, America)$ | $\sim Weapon(y) \vee \sim Sells(West, y, z) \vee \sim Hostile(z)$ | | $\sim Weapon(y) \vee \sim Sells(West, y,$ $\vee \sim Hostile(z) \vee Enemy(Nono, America)$ |

| | | | |
|---|---|---|---|
| $\sim Enemy(Nono, America)$ $\sim Hostile(Nono)$ | $\sim Weapon(y) \vee \sim Sells(West, y, Nono) \vee \sim Hostile(Nono) \vee Enemy(Nono, America)$ | z=Nono | $\sim Weapon(y) \vee \sim Sells(West, y, Nono)$ |
| $Missle(M_i)$ | $\sim Weapon(y) \vee \sim Sells(West, y, Nono)$ | | $\sim Weapon(y) \vee \sim Sells(West, y, Nono) \vee Missle(M_i)$ |
| $\sim Missle(M_i) \vee Weapon(M_i)$ | $\sim Weapon(M_i) \vee \sim Sells(West, M_i, Nono) \vee Missle(M_i)$ | y = Mi | $\sim Sells(West, M_i, Nono)$ |
| $\sim Missle(M_i) \vee \sim Owns(Nono, M_i) \vee Sells(West, M_i, Nono)$ | $\sim Sells(West, M_i, Nono)$ | x =Mi | $\sim Missle(M_i) \vee \sim Owns(Nono, M$ |
| $Owns(Nono, M_i)$ | $\sim Missle(M_i) \vee \sim Owns(Nono, M_i)$ | | $\sim Missle(M_i)$ <u>Contradiction with F6.</u> |

[25 marks]

2. a) $O = f\left(\sum_{i=1}^{n} I_i w_i + w_0\right) = f\left(\sum_{i=0}^{n} I_i w_i\right)$ if we always set $I_0 = 1$ [5 marks]

b) Let $h_{i,k}$ is the output of node $N_{i,k}$

$h_{i,k} = f\left(\sum_{j=0}^{H_{i-1}} h_{i-1,j} w_{i-1,j,k}\right)$ for $i = 1, 2, 3, \ldots, K, K+1$ where $h_{0,k} = I_j$ for k >0, $h_{i-1,0} = 1$

$$O_m = h_{k+1,m} = f\left(\sum_{j=0}^{H_k} w_{k,j,m} h_{k,j}\right)$$

[2.5+2.5 marks]

c) For $f(z) = \frac{1}{1+e^{-z}}$, $f'(z) = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \frac{-1}{(1+e^{-z})^2} \cdot \frac{d}{dz}(1+e^{-z}) = \frac{-1}{(1+e^{-z})^2} \cdot (-e^{-z})$

$$= \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{(1+e^{-z})-1}{1+e^{-z}} = f(z)[1-f(z)]$$

[5 marks]

d) Because $-g'(x)$ is the direction that $g$ decreases most rapidly, but numerically we need to know how large a step size to "move". The learning rate lets us tune a suitable step size. (It is now suitable to think about the effects of different learning rate on the optimization)

[5 marks]

e)

i) For the output neurons,

$$\frac{\partial E}{\partial w_{K,j,k}} = \frac{\partial E}{\partial O_k}\cdot\frac{\partial O_k}{\partial w_{K,j,k}}$$

$$= (O_k - T_k)\cdot f'\left(\sum_{\hat{j}=0}^{H_K} h_{K,\hat{j}} w_{K,\hat{j},k}\right)\cdot h_{K,j} \quad \text{where } h_{K,0} = 1$$

$$= [(O_k - T_k)\cdot h_{K+1,k}\cdot(1 - h_{K+1,k})]\cdot h_{K,j}$$ by part (c) [10 marks]

ii) For the hidden neurons,

$$\frac{\partial E}{\partial h_{i+1,k}} = \sum_{\hat{k}=1}^{H_{i+2}} \frac{\partial E}{\partial h_{i+2,\hat{k}}}\cdot\frac{\partial h_{i+2,\hat{k}}}{\partial h_{i+1,k}}$$ (by Multivariate Chain Rule)

$$= \sum_{\hat{k}=1}^{H_{i+2}} \frac{\partial E}{\partial h_{i+2,\hat{k}}}\cdot h_{i+2,\hat{k}}\cdot(1 - h_{i+2,\hat{k}})\cdot w_{i+1,k,\hat{k}}$$

Let $\Delta_{i,k} = \frac{\partial E}{\partial h_{i,k}}\cdot h_{i,k}\cdot(1 - h_{i,k})$

$$\frac{\partial E}{\partial h_{i+1,k}} = \sum_{\hat{k}=1}^{H_{i+2}} \Delta_{i+2,\hat{k}}\cdot w_{i+1,k,\hat{k}}$$ [10 marks]

iii) For the hidden neurons,

$$\frac{\partial E}{\partial w_{i,j,k}} = \frac{\partial E}{\partial h_{i+1,k}}\cdot\frac{\partial h_{i+1,k}}{\partial w_{i,j,k}} = \frac{\partial E}{\partial h_{i+1,k}}\cdot h_{i+1,k}\cdot(1 - h_{i+1,k})\cdot h_{i,j} = \Delta_{i+1,k}\cdot h_{i,j}$$ [10 marks]

iv) Calculate $\frac{\partial E}{\partial w_{K,j,k}}$'s and $\frac{\partial E}{\partial O_k}$'s by part (ei);

Update $w'_{K,j,k} \leftarrow w_{K,j,k} - \alpha\frac{\partial E}{\partial w_{K,j,k}}$;

For $i$ = $K$-1 downto 0

Calculate $\frac{\partial E}{\partial h_{i+1,k}}$'s by part (eii);

Calculate $\frac{\partial E}{\partial w_{i,j,k}}$'s by part (eiii);

Update $w'_{i,j,k} \leftarrow w_{i,j,k} - \alpha\frac{\partial E}{\partial w_{i,j,k}}$; [5 marks]