1. Overweight

$H = -\left(\frac{4}{6} \times \log_2\left(\frac{4}{6}\right) + \frac{1}{6} \times \log_2\left(\frac{2}{6}\right)\right) \approx 0.918$

1: Yes/boy
2: No/Girl.

Gender

$E_1 = -\left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) + \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$

$E_2 = -\left(1 \times \log_2(1) + 0 \times \log_2(0)\right) = 0.$

$E(\text{gender}) = \frac{4}{6} \times E_1 + \frac{2}{6} \times E_2 = \frac{2}{3}$ $\quad g = 0.918 - \frac{2}{3} = 0.251$

Hyper...

$E_1 \approx 0.$

$E_2 = -\left(\frac{1}{3} \times \log\left(\frac{1}{3}\right) + \frac{2}{3} \times \log\left(\frac{2}{3}\right)\right) \approx 0.918$

$E(H) = \frac{1}{2} \times E_1 + \frac{1}{2} \times E_2 = \frac{0.918}{2}$ $\quad g = 0.459.$

Unhealthy:

$E_1 = -\left(\frac{3}{4} \times \log\left(\frac{3}{4}\right) + \frac{1}{4} \times \log\left(\frac{1}{4}\right)\right) = 0.811$

$E_2 = -\left(\frac{1}{2} \times \log\left(\frac{1}{2}\right) + \frac{1}{2} \times \log\left(\frac{1}{2}\right)\right) = -1$

$E(UD) = \frac{4}{6} \times E_1 + \frac{2}{6} \times E_2 = 0.874$ $\quad g = 0.044.$

Exercise:

$E_1 = -\left(\frac{1}{4} \times \log\left(\frac{2}{4}\right) + \frac{2}{4} \times \log\left(\frac{2}{4}\right)\right) = 1$

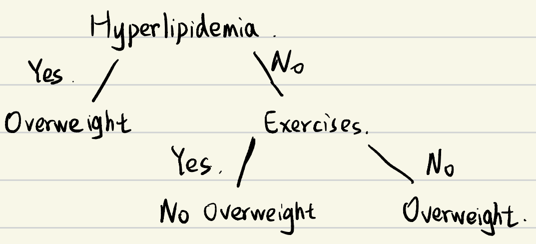
$E_2 = 0.$

$E(E) = \frac{4}{6} \times E_1 + \frac{2}{6} \times E_2 = \frac{2}{3}$ $\quad g = 0.251$

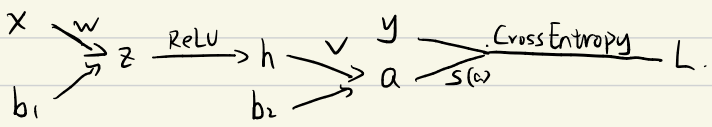
=> Hyperlipidemia is the root.

With 5o observations we can find Exercise is the node to "No" of Hyperlipidemia.

=> The tree:



1.2.



$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial V} = u_2^T h^T = (p-y)^T h^T$ $\quad \Rightarrow \psi_1(u_2, h) = \left[(p-y)^T h^T\right]_{1,:}$

$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial b_2} = u_2^T = (p-y)^T$ $\quad \Rightarrow \psi_2(u_2) = p-y$

$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial W} = \left[(V^T u_2) \odot H(z)\right] x^T$ $\quad \Rightarrow \psi_3(u_1, x) = \left[\left[(V^T u_2) \odot H(z)\right] x^T\right]_{1,:}$

$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial b_1} = u_1^T = \left[(V^T u_2) \odot H(z)\right]^T$ $\quad \Rightarrow \psi_4(u_1) = (V^T u_2) \odot H(z)$

$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial x} = \left[(V^T u_2) \odot H(z)\right] W$ $\quad \Rightarrow \psi_5(W, u_1) = W^T\left[(V^T u_2) \odot H(z)\right]$

1.3.

The neural network has no hidden layer => linear model.

=> Output: $z_k = \sum_{i=1}^{n} w_i x_i + b_i$

Using softmax to change it to probability:

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

For Multinomial logistic Model.

$$P(Y=k|x) = \frac{e^{\beta_k \cdot x_i}}{\sum_{j=1}^{K} e^{z_j}}.$$

They are identical, for MLM, it is a NN with $b_i = 0,\ i = 1 \ldots k$.

1.4. (a) Conv5 (10): $(32-5+2\times2)+1 = 32\times32\times10$.

Maxpool2 : $(32-2)\div2+1 = 16\times16\times10$.

Conv5 (10) : $(16-5+2\times2)+1 = 16\times16\times10$.

Maxpool2 : $(16-2)\div2+1 = 8\times8\times10$.

FC10 : 10.

(b) Conv5 (10)(1) : $(5\times5\times3+1)\times10 = 760$.

Conv5 (10)(2) : $(5\times5\times10+1)\times10 = 2510$.

FC10 : $(8\times8\times10)\times10+10 = 6410$.

1.5.

1.5(a)

When some neurons are randomly dropped out during training, the network cannot rely on some particular neurons to always be present, instead, they need to learn to make use of a variety of different combinations of neurons.
This helps to prevent the network from becoming too specialized to the training data and encourages it to learn more generalizable features.

1.5(b)

The probability of a neuron to be remained is $1-p$, and with the expetation should maintain the same, we need to divide every remained activations by $\frac{1}{1-p}$, thus the expetation would remain the same.

(c) $$\frac{\partial z}{\partial h^{(l)}} = \frac{\partial z}{\partial h^{(l+1)}} \cdot \frac{\partial h^{(l+1)}}{\partial h^{(l)}}$$

$$= \frac{\partial z}{\partial h^{(l+1)}} \odot \text{mask}.$$

1.6. (a) $f(x) = \sigma(0.5x_1 - 0.1x_2)$

Positive: $f(5,5) = \sigma(2) = 0.8808$.

$f(3,8) = \sigma(0.7) = 0.6682$.

$f(-1,8) = \sigma(-1.3) = 0.2142$.

$f(5,4) = \sigma(2.1) = 0.8909$

$f(-1,-2) = \sigma(-0.3) = 0.4256$.

Negative: $f(-5,1) = \sigma(-2.6) = 0.067$

$f(2,-2) = \sigma(1.2) = 0.7685$.

$f(-1,1) = \sigma(-0.6) = 0.3544$

$f(5,-10) = \sigma(3.5) = 0.9707$

$f(-10,-9) = \sigma(-4.1) = 0.016$.

(b) threhold $= 0.5$ $\Rightarrow$ Positive $= [1,1,0,1,0]$ $\Rightarrow$ $TP = 3$ $FN = 2$

Negative $= [0,1,0,1,0]$ $FP = 2$ $TN = 3$.

Accuracy $= 0.5$.

Precision $= 0.6$.

Recall $= 0.6$.

$F1 = \frac{2 \times 0.6 \times 0.6}{0.6 + 0.6} = 0.6$.

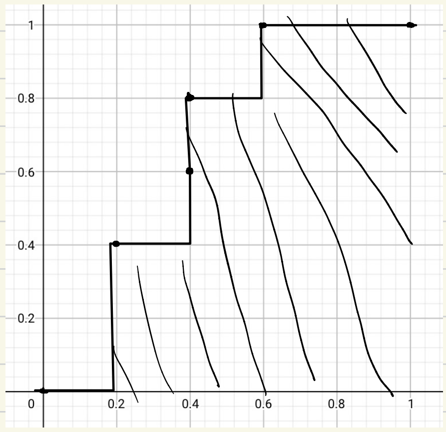
Confusion Matrix:

	$\hat{P}$	$\hat{N}$
P	3	2
N	2	3.

(c) Threshold $= 0.$ $TPR = 1$, $FPR = 1$

0.2	$TPR = 1$	$FPR = 0.6$.
0.4	$TPR = 0.8$	$FPR = 0.4$
0.6	$TPR = 0.6$	$FPR = 0.4$
0.8	$TPR = 0.4$	$FPR = 0.2$
1.0	$TPR = 0$	$FPR = 0$.

TPR



FPR

(d) $AUC = 0.2 \times 0.4 + 0.2 \times 0.8 + 0.4 \times 1 = 0.64$