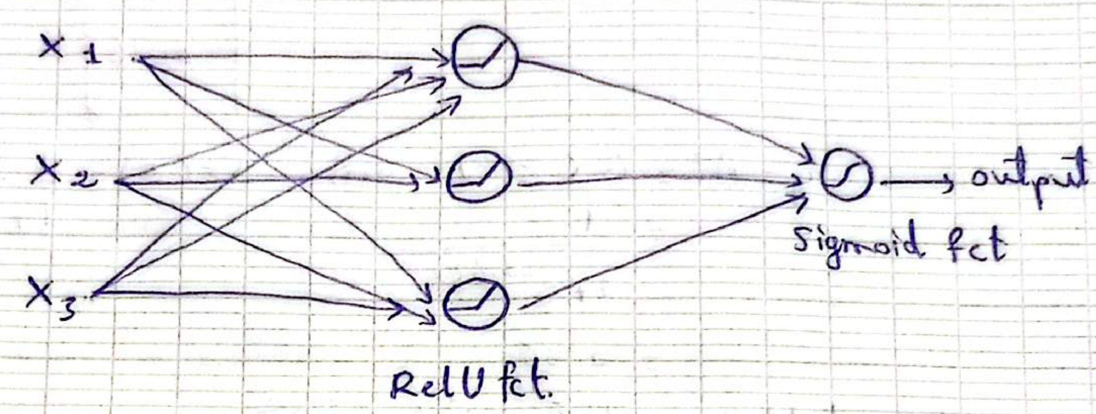
Final 2025-2026 Solution:

Ex 1: 1) Input layer (3 input features) — Hidden layer (3 neurons) — Output layer (1 neuron)



• Neural Network Architecture.

2) Hidden layer:

$$Z^1 = W^1 X + b^1$$

$$= \begin{bmatrix} 0{,}2 & -0{,}5 & 0{,}3 \\ -0{,}4 & 0{,}1 & 0{,}6 \\ 0{,}7 & -0{,}2 & 0{,}1 \end{bmatrix} \begin{bmatrix} 0{,}8 \\ -0{,}4 \\ 1{,}2 \end{bmatrix} + \begin{bmatrix} 0{,}1 \\ -0{,}2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0{,}82 \\ 0{,}16 \\ 0{,}76 \end{bmatrix}$$

$A^1 = G(Z^1)$ with $G(Z) = \max(0, Z)$

$$= \begin{bmatrix} \max(0, 0{,}82) \\ \max(0, 0{,}16) \\ \max(0, 0{,}76) \end{bmatrix} = \begin{bmatrix} 0{,}82 \\ 0{,}16 \\ 0{,}76 \end{bmatrix}$$

3) Output layer:

$$Z^2 = W^2 A^1 + b^2$$

$$= \begin{bmatrix} 0.5 & -0.7 & 0.9 \end{bmatrix} \begin{bmatrix} 0.82 \\ 0.16 \\ 0.76 \end{bmatrix} + [-0.3] = 0.682$$

$$\hat{y} = A^2 = G(Z^2) \text{ with } G(z) = \frac{1}{1+e^{-z}}$$

$$= \frac{1}{1+e^{-0.682}}$$

$$= 0.66$$

4) Using $MSE = \sum (y-\hat{y})^2$:

error $=$ MSE $= 1 - 0.66 = 0.34$

The loss value is the same as the error.

5) Find gradients:

- $$\frac{dL}{dw^2} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dZ^2} \cdot \frac{dZ^2}{dw^2}$$

with: $$\frac{d\hat{y}}{dZ^2} = (\text{Sigmoid})' = \hat{y}(1-\hat{y})$$

$$\frac{dL}{d\hat{y}} = -2(y-\hat{y}) \quad , \quad \frac{dZ^2}{dw^2} = A^1$$

$$\Rightarrow \frac{dL}{dw^2} = -2\hat{y}.(1-\hat{y})(y-\hat{y}).A^1$$

- $$\frac{dL}{dw^1} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dZ^2} \cdot \frac{dZ^2}{dA^1} \cdot \frac{dA^1}{dZ^1} \cdot \frac{dZ^1}{dw^1}$$

with $$\frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dZ^2} = -2\hat{y}(1-\hat{y})(y-\hat{y})$$

$$\frac{dz^2}{dA^1} = w^2 \quad , \quad \frac{dA^1}{dz^1} = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{else} \end{cases} \quad , \quad \frac{dz^1}{dw^1} = X$$

$$\Rightarrow \frac{dl}{dw^1} = -2\hat{y}(1-\hat{y})(y-\hat{y}) \cdot w^2 \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot X$$

6) Using gradient descent: $w = w - \eta \frac{dl}{dw}$ with $\eta = 0.01$

- $w^1 = w^1 - \eta \frac{dl}{dw^1}$ where $\frac{dl}{dw^1} = [-0.06 \quad -0.04 \quad -0.16]$

$$= \begin{bmatrix} 0.2 & -0.5 & 0.3 \\ -0.4 & 0.1 & 0.6 \\ 0.7 & -0.2 & 0.1 \end{bmatrix} - (0.01) \begin{bmatrix} -0.06 & -0.04 & -0.16 \\ -0.06 & -0.04 & -0.16 \\ -0.06 & -0.04 & -0.16 \end{bmatrix}$$

$$= \begin{bmatrix} 0.2006 & -0.4996 & 0.3016 \\ -0.3994 & 0.1004 & 0.6016 \\ 0.7006 & -0.1996 & 0.1016 \end{bmatrix}$$

- $w^2 = w^2 - \eta \frac{dL}{dw^2}$ where $\frac{dl}{dw^2} = [-0.123 \quad -0.024 \quad -0.114]$

$$= [0.5 \quad -0.7 \quad 0.9] - (0.01)[-0.123 \quad -0.024 \quad -0.114]$$

$$= [0.50123 \quad -0.69976 \quad 0.90114]$$

Ex 3:

1) - Learning phase:

$P(play = Yes) = 5/8$

$P(play = No) = 3/8$

- Conditional probabilities of features:

Outlook	Yes	No
Sunny	1/5	2/3
overcast	2/5	0/3
Rainy	2/5	1/3

Temperature	Yes	No
Hot	1/5	2/3
Mild	2/5	0/3
Cool	2/5	1/3

- Testing phase:

- S1: $P(Yes/S1) = P(sunny/Yes) \times P(Mild/Yes) \times P(Yes)$

$= \frac{1}{5} \times \frac{2}{5} \times \frac{5}{8} = 0.05$

$P(No/S1) = P(Sunny/No) \times P(Mild/No) \times P(No)$

$= \frac{2}{3} \times 0 \times \frac{3}{8} = 0$

$P(Yes/S1) > P(No/S1) \Rightarrow$ S1 class : play = Yes

- S2: $P(Yes/S2) = \frac{1}{5} \times \frac{2}{5} \times \frac{5}{8} = 0.05$

$P(No/S2) = \frac{2}{3} \times \frac{1}{3} \times \frac{3}{8} = 0.083$

$\Rightarrow$ S2: play = No

- S3: $P(Yes/S3) = \frac{1}{5} \times \frac{1}{5} \times \frac{5}{8} = 0.025$

$P(No/S3) = \frac{2}{3} \times \frac{2}{3} \times \frac{3}{8} = 0.16$

$\Rightarrow$ S3: play = No

- $S4: P(Yes/S4) = \frac{2}{5} \times \frac{2}{5} \times \frac{5}{8} = 0.1$

  $P(No/S4) = \frac{1}{3} \times 0 \times \frac{3}{8} = 0$

$\rightarrow S4: play = Yes$

- $S5: P(Yes/S5) = \frac{2}{5} \times \frac{1}{5} \times \frac{5}{8}$

  $P(No/S5) = \frac{0}{3} \times \dots = 0$

$\Rightarrow S5: play = Yes$

- $S6: P(Yes/S6) = \frac{2}{5} \times \frac{2}{5} \times \frac{5}{8}$

  $P(No/S6) = \frac{0}{3} \times \_ = 0$

$\Rightarrow S6: play = Yes$

2) a) confusion matrix:

TP

		Predicted: Yes	Predicted: No	
actual	Yes	2	1	→ FN
	No	2	1	→ TN

FP

b) $Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{2+1}{2+1+2+1} = 0,5 = 50\%$

$Precision = \frac{TP}{TP + FP} = \frac{2}{2+2} = 0.5$

$recall = \frac{TP}{TP + FN} = \frac{2}{2+1} = 0.67$

$F1\text{-score} = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.5 \times 0.67}{0.5 + 0.67} = 0.57$