Q 1

(a) Derive the gradient of the BCE loss function with respect to the weights of the output layer

$$L(\boldsymbol{W}) = -[\boldsymbol{y} * \log(\hat{\boldsymbol{y}}) + (1 - \mathbf{y}) * \log(1 - \hat{\boldsymbol{y}})] \quad (1)$$

$$\hat{\boldsymbol{y}} = \sigma(z) = \frac{1}{1 + e^{-\boldsymbol{z}}} \quad (2)$$

$$\frac{\partial L(\boldsymbol{W})}{\partial \boldsymbol{W}} = \frac{\partial L(\boldsymbol{W})}{\partial \hat{\boldsymbol{y}}} \frac{\partial \hat{\boldsymbol{y}}}{\partial \boldsymbol{z}} \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{W}} \quad (4)$$

$$\frac{\partial L(\boldsymbol{W})}{\partial \hat{\boldsymbol{y}}} = -\left(\frac{\boldsymbol{y}}{\hat{\boldsymbol{y}}} - \frac{1 - \boldsymbol{y}}{1 - \hat{\boldsymbol{y}}}\right) = \frac{\hat{\boldsymbol{y}} - \boldsymbol{y}}{\hat{\boldsymbol{y}}(1 - \hat{\boldsymbol{y}})} \quad (5)$$

$$\frac{\partial \hat{\boldsymbol{y}}}{\partial \boldsymbol{z}} = \frac{\partial}{\partial \boldsymbol{z}}\left[\frac{1}{1 + e^{-\boldsymbol{z}}}\right] = \frac{\partial}{\partial \boldsymbol{z}}(1 + e^{-\boldsymbol{z}})^{-1} = \quad (6)$$

omitting....

$$= \frac{1}{1 + e^{-\boldsymbol{z}}}\left(1 - \frac{1}{1 + e^{-\boldsymbol{z}}}\right) = \hat{\boldsymbol{y}}(1 - \hat{\boldsymbol{y}}) \quad (10)$$

$$\frac{\partial z}{\partial W} = X \tag{11}$$

$$\frac{\partial L(W)}{\partial W} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})X = (\hat{y} - y)X \tag{12}$$

where X is output weights (vector)

(b)Explain how the gradient is influenced by the magnitude of the predictions (close to 0 or 1).

	$\hat{y}$ Close to 1	$\hat{y}$ Close to 0
y=1	the gradient is small	large negative gradient
y=0	large positive gradient	the gradient is small

Q 2

(a) Calculate the accuracy, precision, recall, and F1-score for each class.

Precision_A = $\frac{50}{50+7+4}$ = 0.82; Recall_A = $\frac{50}{50+5+10}$ = 0.77; F1_A= $\frac{2*0.82*0.77}{0.82+0.77}$ =0.79

Precision_B = $\frac{60}{60+5+6}$ = 0.85; Recall_B = $\frac{60}{60+7+8}$ = 0.80; F1_B= $\frac{2*0.85*0.80}{0.85+0.80}$ =0.82

Precision_C = $\frac{80}{80+8+10}$ = 0.82; Recall_C= $\frac{80}{80+6+4}$ = 0.89; F1_C= $\frac{2*0.82*0.89}{0.82+0.89}$ =0.85

(b)

Accuracy = $\frac{50+60+80}{50+5+10+7+60+8+4+6+80}$ = 0.83

Class	F1	Support	Support Proportion	weighted average F1-score
A	0.74	65	0.28	0.79*0.28 + 0.82*0.32 + 0.85*0.39=**0.82**
B	0.82	75	0.32	
C	0.85	90	0.39	