2.1

Q_1 $x \cdot W$

$$[\quad]_{1\times 4} \begin{bmatrix} & \\ & \\ & \\ & \end{bmatrix}_{4\times 2} = [\quad]_{1\times 2}$$

The network will have 2 classes

Q_2 $p = x \cdot W = \begin{bmatrix} -2 & 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 2 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -5 \end{bmatrix}$

$q = \begin{bmatrix} \sigma(6) & \sigma(-5) \end{bmatrix} = \begin{bmatrix} 0.9975 & 0.006693 \end{bmatrix}$

$\frac{1}{1+e^{-6}}$ $\qquad$ $\frac{1}{1+e^{5}}$

$L = \Sigma q = \sigma(6) + \sigma(-5) = 1.0042$

Q_3 $L = q_1 + q_2$ $\rightarrow$ $\frac{\partial L}{\partial q} = \begin{bmatrix} 1 & 1 \end{bmatrix}$

$q_1 = \frac{1}{1+e^{p_1}}$ $\qquad$ $q_2 = \frac{1}{1+e^{p_2}}$ $\rightarrow$ $\frac{\partial q_1}{\partial p_1} = \sigma(p_1) \cdot (1 - \sigma(p_1)) = 0.0025$

$\frac{\partial q_2}{\partial p_2} = \sigma(p_2) \cdot (1 - \sigma(p_2)) = 0.006648$

$p = x \cdot W$ $\rightarrow$ $\frac{\partial p}{\partial x} = W^T$

$\frac{\partial p}{\partial W} = x^T$

Q4

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial p} \cdot \frac{\partial p}{\partial x} = \underset{1\times 2}{\begin{bmatrix} 1 & 1 \end{bmatrix}} \begin{bmatrix} 0.0025 & 0 \\ 0 & 0.006648 \end{bmatrix} \begin{bmatrix} -1 & 2 & 1 & 0 \\ 3 & -1 & 1 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 0.0174 & -0.00165 & 0.009148 & 0.006648 \end{bmatrix}$$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial q} \frac{\partial q}{\partial p} \cdot \frac{\partial p}{\partial W} = \underset{2\times 1}{\left(\underset{1\times 2}{\begin{bmatrix} 1 & 1 \end{bmatrix}} \underset{2\times 2}{\begin{bmatrix} 0.0025 & 0 \\ 0 & 0.006648 \end{bmatrix}} \right)^T} \underset{1\times 4}{\begin{bmatrix} -2 & 1 & 2 & 0 \end{bmatrix}} =$$

$$\underset{2\times 1}{\begin{bmatrix} 0.0025 \\ 0.006648 \end{bmatrix}} \underset{1\times 4}{\begin{bmatrix} -2 & 1 & 2 & 0 \end{bmatrix}} = \begin{bmatrix} -0.01 & 0.0025 & 0.005 & 0 \\ -0.026 & 0.006648 & 0.0132 & 0 \end{bmatrix}$$

2.2

1. This is likely due to overfitting, making the model tailored to what it was trained with. The solution is to either decrease the amount of training data. Data augmentation and a wider amount of data would also aid in reducing the effects of overfitting.
2. False-Positive errors occur when the wrong data is marked as positive. This could be an issue in autonomous driving, where a shadow could be detected as an object, causing the vehicle to abruptly stop, which may lead to an accident.

False-Negative errors occurs when data is incorrectly marked as negative. This can have horrible consequences in cancer detection, where a patient is told they don't have cancer when they actually do.

3.2

1)

$$\begin{matrix} 6 & 3 & \boxed{10} & \boxed{32} & 0 & 1 \\ 5 & -1 & \boxed{0} & \boxed{11} & -2 & 2 \\ \boxed{-32} & \boxed{-1} & 7 & 1 & -2 & 3 \\ \boxed{-1} & \boxed{-1} & 0 & -1 & -2 & 4 \\ -1 & -1 & 0 & -1 & -2 & 5 \end{matrix} \rightarrow \begin{matrix} 6 & 32 & 2 \\ -1 & 7 & 4 \end{matrix}$$

2)

$$\begin{matrix} \boxed{6} & \boxed{3} & 10 & 32 & 0 & 1 \\ \boxed{5} & \boxed{-1} & 0 & 11 & -2 & 2 \\ -32 & -1 & \boxed{7} & \boxed{1} & -2 & 3 \\ -1 & -1 & \boxed{0} & \boxed{-1} & -2 & 4 \\ -1 & -1 & 0 & -1 & -2 & 5 \end{matrix} \rightarrow \begin{matrix} 3.25 & 13.25 & 0.25 \\ -8.75 & 1.75 & 0.75 \end{matrix}$$

3) No. A 4D kernel will not work with a 2D tensor

4) $[\ ]_{10\times10} \rightarrow [\ ]\ 10\times10 \rightarrow [\ ]_{5\times5} \rightarrow$ 3 outputs

3x3 filter

$\underbrace{2}_{\text{padding}}+10-3+1 = 10$

2x2 max pool

$H=W=\dfrac{10-2}{2}+1=5$

Weights: $3\times3 + 5\cdot5\cdot3 = 84$ weights

5) Valid = No padding. 1x1 output since input size == filter

$$X_0 \cdot W_0 = -1\cdot 2 + -1\cdot 2 + 1\cdot 2 = -2$$

$$X_1 \cdot W_1 = 1 + 1 + 1 + 1 = 4$$

$$X_2 \cdot W_2 = 2\cdot 1 = 2$$

$$\text{Output} = -2 + 4 + 2 + 1 = \underline{5}$$

6) Same = padding to ensure same output size

$$x_0 = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \cdot W_0 \qquad X_1 = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \cdot W_1 \qquad X_2 = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \cdot W_2$$

filter