2025Fall_CS526_HW2

xc166

October 2025

1 Question 2

(a): Mean of Residuals is Zero

To Prove:

In simple linear regression, the mean of the residuals e_i is always zero:

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = 0$$

Proof:

Step 1: Define the residual

The residual for observation i is:

$$e_i = Y_i - \hat{Y}_i$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is the predicted value.

Step 2: Sum the residuals

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)$$

Substitute $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]$$

$$= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 X_i$$

$$= \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i$$

Step 3: Use the normal equation

In simple linear regression, the least squares estimates satisfy the **first normal equation**:

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

This is equivalent to:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i$$

Step 4: Substitute back

From Step 2, we have:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i$$

Using the normal equation from Step 3:

$$\sum_{i=1}^{n} e_i = 0$$

Step 5: Calculate the mean

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n} \cdot 0 = 0$$

(b): Residuals are Orthogonal to Predictor

To Prove:

In simple linear regression, the residuals e_i are orthogonal to the predictor variable X_i:

$$\sum_{i=1}^{n} X_i e_i = 0$$

Proof:

Step 1: Define the residual

The residual for observation i is:

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$$

Step 2: Compute the dot product

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

Expand:

$$= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \hat{\beta}_0 - \sum_{i=1}^{n} X_i \hat{\beta}_1 X_i$$

$$= \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

Step 3: Use the second normal equation

In simple linear regression, the least squares estimates satisfy the **second normal equation**:

$$\sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

This is obtained by taking the partial derivative of $\sum e_i^2$ with respect to $\hat{\beta}_1$ and setting it to zero:

$$\frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = -2 \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

This gives us:

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

Step 4: Substitute back

From Step 2:

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

Using the second normal equation from Step 3:

$$\sum_{i=1}^{n} X_i e_i = 0$$

(c): Residuals Uncorrelated with Predicted Response

To Prove:

In simple linear regression, the residuals are uncorrelated with the predicted responses:

$$\mathrm{Cov}(e, \hat{Y}) = \frac{1}{n} \sum_{i=1}^{n} (e_i - \bar{e})(\hat{Y}_i - \bar{\hat{Y}}) = 0$$

Proof:

Step 1: Simplify using result from part (a)

From part (a), we know that $\bar{e} = 0$.
Therefore:

$$\text{Cov}(e, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(e_i - 0)(\hat{Y}_i - \bar{\hat{Y}})$$

$$= \frac{1}{n}\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{\hat{Y}})$$

Step 2: Expand the predicted value

Recall that $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.
Also, $\bar{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$
Therefore:

$$\hat{Y}_i - \bar{\hat{Y}} = (\hat{\beta}_0 + \hat{\beta}_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1\bar{X})$$

$$= \hat{\beta}_1(X_i - \bar{X})$$

Step 3: Substitute into covariance

$$\text{Cov}(e, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} e_i \cdot \hat{\beta}_1(X_i - \bar{X})$$

$$= \frac{\hat{\beta}_1}{n}\sum_{i=1}^{n} e_i(X_i - \bar{X})$$

$$= \frac{\hat{\beta}_1}{n}\left[\sum_{i=1}^{n} e_i X_i - \sum_{i=1}^{n} e_i\bar{X}\right]$$

$$= \frac{\hat{\beta}_1}{n}\left[\sum_{i=1}^{n} e_i X_i - \bar{X}\sum_{i=1}^{n} e_i\right]$$

Step 4: Apply results from parts (a) and (b)

From part (a): $\sum_{i=1}^{n} e_i = 0$
From part (b): $\sum_{i=1}^{n} X_i e_i = 0$
Therefore:

$$\text{Cov}(e, \hat{Y}) = \frac{\hat{\beta}_1}{n}[0 - \bar{X} \cdot 0] = 0$$

(d): Mean of Predicted Equals Mean of Observed

To Prove:

In simple linear regression, the mean of the predicted responses equals the mean of the observed responses:

$$\bar{\hat{Y}} = \bar{Y}$$

Proof:

Step 1: Express the relationship between Y, $\hat{Y}$, and e

For each observation:

$$Y_i = \hat{Y}_i + e_i$$

Step 2: Sum both sides

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i + \sum_{i=1}^{n} e_i$$

Step 3: Apply result from part (a)

From part (a), we know that $\sum_{i=1}^{n} e_i = 0$.
Therefore:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i + 0$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

Step 4: Divide by n to get means

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i$$

$$\bar{Y} = \bar{\hat{Y}}$$

(e): Proof that $R^2 = \frac{ESS}{TSS}$

To Prove:

Starting from the definition $R^2 = 1 - \frac{RSS}{TSS}$, show that:

$$R^2 = \frac{ESS}{TSS}$$

where:

- $RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ (Residual Sum of Squares)
- $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ (Total Sum of Squares)
- $ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ (Explained Sum of Squares)

Proof:

Step 1: Decompose the total deviation

For each observation, we can write:

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + e_i$$

Step 2: Square both sides and sum

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}[(\hat{Y}_i - \bar{Y}) + e_i]^2$$

Expand the right side:

$$= \sum_{i=1}^{n}[(\hat{Y}_i - \bar{Y})^2 + 2(\hat{Y}_i - \bar{Y})e_i + e_i^2]$$

$$= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})e_i + \sum_{i=1}^{n}e_i^2$$

Step 3: Show the cross-product term equals zero

We need to show that $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})e_i = 0$.

From part (d), we know $\bar{\hat{Y}} = \bar{Y}$.

Therefore:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})e_i = \sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})e_i$$

From part (c), we proved that:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})e_i = n \cdot \text{Cov}(e, \hat{Y}) = n \cdot 0 = 0$$

Step 4: Establish TSS = ESS + RSS

From Step 2 and Step 3:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}e_i^2$$

$$TSS = ESS + RSS$$

Step 5: Derive $R^2 = \frac{ESS}{TSS}$

From the definition:

$$R^2 = 1 - \frac{RSS}{TSS}$$

From Step 4, we have $RSS = TSS - ESS$, so:

$$R^2 = 1 - \frac{TSS - ESS}{TSS}$$

$$= 1 - \frac{TSS}{TSS} + \frac{ESS}{TSS}$$

$$= 1 - 1 + \frac{ESS}{TSS}$$

$$= \frac{ESS}{TSS}$$

(f): Proof that $R^2 = r_{XY}^2$

To Prove:

In simple linear regression with Y as the response and X as the predictor, the R^2 statistic equals the square of the correlation coefficient:

$$R^2 = r_{XY}^2$$

where:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Proof:

Step 1: Recall the formula for $\hat{\beta}_1$

In simple linear regression, the slope is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Step 2: Express R^2 using part (e)

From part (e), we know:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Step 3: Simplify the numerator (ESS)

From part (c), we showed that:

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_1(X_i - \bar{X})$$

Therefore:

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}[\hat{\beta}_1(X_i - \bar{X})]^2$$

$$= \hat{\beta}_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Step 4: Substitute into R^2

$$R^2 = \frac{\hat{\beta}_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

Step 5: Substitute the formula for $\hat{\beta}_1$
From Step 1:

$$\hat{\beta}_1^2 = \left[\frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right]^2$$

$$= \frac{[\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})]^2}{[\sum_{i=1}^{n} (X_i - \bar{X})^2]^2}$$

Substitute into Step 4:

$$R^2 = \frac{[\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})]^2}{[\sum_{i=1}^{n} (X_i - \bar{X})^2]^2} \cdot \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

$$= \frac{[\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 \cdot \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

Step 6: Recognize the correlation coefficient
The expression above is exactly:

$$R^2 = \left[\frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \right]^2 = r_{XY}^2$$

2 Question 3: Logistic Regression Manual Calculations

Given Data:

Sample	Sugar Intake (x)	Condition (y)
1	30	0
2	50	0
3	70	1
4	90	1

Initial parameters: $\theta_0 = 0$, $\theta_1 = 0$

(a) Calculate $h_\theta(x)$ for each training example

The sigmoid function is:

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 \cdot x)}}$$

With $\theta_0 = 0$ and $\theta_1 = 0$:

$$h_\theta(x) = \frac{1}{1 + e^{-(0 + 0 \cdot x)}} = \frac{1}{1 + e^0} = \frac{1}{1 + 1} = \frac{1}{2} = 0.5$$

Calculations:
Sample 1: $x^{(1)} = 30$

$$h_\theta(30) = 0.5$$

Sample 2: $x^{(2)} = 50$

$$h_\theta(50) = 0.5$$

Sample 3: $x^{(3)} = 70$

$$h_\theta(70) = 0.5$$

Sample 4: $x^{(4)} = 90$

$$h_\theta(90) = 0.5$$

Result: All samples have $h_\theta(x^{(i)}) = 0.5$ because the initial parameters are both zero.

(b) Calculate the log-likelihood $\ell(\theta)$

The log-likelihood function is:

$$\ell(\theta) = \sum_{i=1}^{4} \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

Since $h_\theta(x^{(i)}) = 0.5$ for all samples:

- $\log(0.5) = \log(1/2) = -\log(2) \approx -0.6931$
- $\log(1 - 0.5) = \log(0.5) = -\log(2) \approx -0.6931$

Calculations by sample:
Sample 1: $x^{(1)} = 30, y^{(1)} = 0$

$$\ell_1 = 0 \cdot \log(0.5) + (1 - 0) \cdot \log(0.5) = \log(0.5) = -\log(2)$$

Sample 2: $x^{(2)} = 50, y^{(2)} = 0$

$$\ell_2 = 0 \cdot \log(0.5) + (1 - 0) \cdot \log(0.5) = \log(0.5) = -\log(2)$$

Sample 3: $x^{(3)} = 70, y^{(3)} = 1$

$$\ell_3 = 1 \cdot \log(0.5) + (1 - 1) \cdot \log(0.5) = \log(0.5) = -\log(2)$$

Sample 4: $x^{(4)} = 90, y^{(4)} = 1$

$$\ell_4 = 1 \cdot \log(0.5) + (1-1) \cdot \log(0.5) = \log(0.5) = -\log(2)$$

Total log-likelihood:

$$\ell(\theta) = \ell_1 + \ell_2 + \ell_3 + \ell_4 = 4 \cdot (-\log(2)) = -4\log(2)$$

$$\ell(\theta) = -4 \times 0.6931 \approx -2.7726$$

(c) First iteration of gradient ascent

Gradient Formulas:

$$\frac{\partial \ell(\theta)}{\partial \theta_0} = \sum_{i=1}^{4} (y^{(i)} - h_\theta(x^{(i)}))$$

$$\frac{\partial \ell(\theta)}{\partial \theta_1} = \sum_{i=1}^{4} (y^{(i)} - h_\theta(x^{(i)})) \cdot x^{(i)}$$

Calculate gradients:
For θ_0:

$$\frac{\partial \ell}{\partial \theta_0} = (0-0.5) + (0-0.5) + (1-0.5) + (1-0.5)$$

$$= -0.5 - 0.5 + 0.5 + 0.5 = 0$$

For θ_1:

$$\frac{\partial \ell}{\partial \theta_1} = (0-0.5) \cdot 30 + (0-0.5) \cdot 50 + (1-0.5) \cdot 70 + (1-0.5) \cdot 90$$

$$= (-0.5)(30) + (-0.5)(50) + (0.5)(70) + (0.5)(90)$$

$$= -15 - 25 + 35 + 45 = 40$$

Update parameters with $\alpha = 0.01$:
Update rule:

$$\theta_j^{\text{new}} = \theta_j^{\text{old}} + \alpha \frac{\partial \ell}{\partial \theta_j}$$

For θ_0:

$$\theta_0^{\text{new}} = 0 + 0.01 \times 0 = 0$$

For θ_1:

$$\theta_1^{\text{new}} = 0 + 0.01 \times 40 = 0.4$$

Updated parameters after first iteration:

$$\theta_0 = 0, \quad \theta_1 = 0.4$$

Explanation of the process:
Gradient ascent is used to maximize the log-likelihood:

1. Start with initial parameters
2. Calculate the gradient (direction of steepest increase)
3. Update parameters in the direction of the gradient: $\theta_j := \theta_j + \alpha \frac{\partial \ell}{\partial \theta_j}$
4. Repeat until convergence

The gradient tells us that increasing θ_1 will increase the log-likelihood, which makes sense because higher sugar intake is associated with higher probability of developing the condition.

3 Question 4: Multiclass Perceptron Manual Calculations

Given Data:

Training samples:

Sample	x_1 (Excitement)	x_2 (Budget)	True Label (y)
1	3	100	0
2	8	300	1
3	5	150	2

Initial weight matrix:

$$W = \begin{bmatrix} 0.4 & 0.1 & -0.3 \\ 0.3 & -0.2 & 0.5 \\ -0.1 & 0.3 & 0.2 \end{bmatrix}$$

Row 1 $\rightarrow$ Category 0, Row 2 $\rightarrow$ Category 1, Row 3 $\rightarrow$ Category 2
Column 1 $\rightarrow$ bias, Column 2 $\rightarrow$ x_1 coefficient, Column 3 $\rightarrow$ x_2 coefficient

(a) Update weights for each training sample

Sample 1: $x_1 = 3, x_2 = 100, y = 0$

Step 1: Form input vector (with bias)

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1 \\ 3 \\ 100 \end{bmatrix}$$

Step 2: Calculate scores for each category
Category 0:

$$\text{score}_0 = 0.4 \cdot 1 + 0.1 \cdot 3 + (-0.3) \cdot 100 = 0.4 + 0.3 - 30 = -29.3$$

Category 1:

$$\text{score}_1 = 0.3 \cdot 1 + (-0.2) \cdot 3 + 0.5 \cdot 100 = 0.3 - 0.6 + 50 = 49.7$$

Category 2:

$$\text{score}_2 = (-0.1) \cdot 1 + 0.3 \cdot 3 + 0.2 \cdot 100 = -0.1 + 0.9 + 20 = 20.8$$

Step 3: Predict category

$$\hat{y}^{(1)} = \arg\max(\text{score}_0, \text{score}_1, \text{score}_2) = \arg\max(-29.3, 49.7, 20.8) = 1$$

Step 4: Check if correct
True label: $y^{(1)} = 0$, Predicted: $\hat{y}^{(1)} = 1$
Prediction is WRONG! $\rightarrow$ Update weights
Step 5: Update weights
Update rule:

- $\mathbf{w}_{\text{true}} := \mathbf{w}_{\text{true}} + \mathbf{x}$ (increase true category)
- $\mathbf{w}_{\text{predicted}} := \mathbf{w}_{\text{predicted}} - \mathbf{x}$ (decrease predicted category)

Increase row 0 (true category):

$$\mathbf{w}_0^{\text{new}} = [0.4, 0.1, -0.3] + [1, 3, 100] = [1.4, 3.1, 99.7]$$

Decrease row 1 (predicted category):

$$\mathbf{w}_1^{\text{new}} = [0.3, -0.2, 0.5] - [1, 3, 100] = [-0.7, -3.2, -99.5]$$

Row 2 unchanged:

$$\mathbf{w}_2^{\text{new}} = [-0.1, 0.3, 0.2]$$

Updated weight matrix after Sample 1:

$$W = \begin{bmatrix} 1.4 & 3.1 & 99.7 \\ -0.7 & -3.2 & -99.5 \\ -0.1 & 0.3 & 0.2 \end{bmatrix}$$

Sample 2: $x_1 = 8, x_2 = 300, y = 1$

Step 1: Form input vector

$$\mathbf{x}^{(2)} = \begin{bmatrix} 1 \\ 8 \\ 300 \end{bmatrix}$$

Step 2: Calculate scores (using updated weights)
Category 0:

$$\text{score}_0 = 1.4 \cdot 1 + 3.1 \cdot 8 + 99.7 \cdot 300 = 1.4 + 24.8 + 29910 = 29936.2$$

Category 1:

$$\text{score}_1 = (-0.7) \cdot 1 + (-3.2) \cdot 8 + (-99.5) \cdot 300 = -0.7 - 25.6 - 29850 = -29876.3$$

Category 2:

$$\text{score}_2 = (-0.1) \cdot 1 + 0.3 \cdot 8 + 0.2 \cdot 300 = -0.1 + 2.4 + 60 = 62.3$$

Step 3: Predict category

$$\hat{y}^{(2)} = \arg\max(29936.2, -29876.3, 62.3) = 0$$

Step 4: Check if correct
True label: $y^{(2)} = 1$, Predicted: $\hat{y}^{(2)} = 0$
Prediction is WRONG! $\rightarrow$ Update weights
Step 5: Update weights
Increase row 1 (true category):

$$\mathbf{w}_1^{\text{new}} = [-0.7, -3.2, -99.5] + [1, 8, 300] = [0.3, 4.8, 200.5]$$

Decrease row 0 (predicted category):

$$\mathbf{w}_0^{\text{new}} = [1.4, 3.1, 99.7] - [1, 8, 300] = [0.4, -4.9, -200.3]$$

Row 2 unchanged:

$$\mathbf{w}_2^{\text{new}} = [-0.1, 0.3, 0.2]$$

Updated weight matrix after Sample 2:

$$W = \begin{bmatrix} 0.4 & -4.9 & -200.3 \\ 0.3 & 4.8 & 200.5 \\ -0.1 & 0.3 & 0.2 \end{bmatrix}$$

Sample 3: $x_1 = 5, x_2 = 150, y = 2$

Step 1: Form input vector

$$\mathbf{x}^{(3)} = \begin{bmatrix} 1 \\ 5 \\ 150 \end{bmatrix}$$

Step 2: Calculate scores (using updated weights)
Category 0:

$$\text{score}_0 = 0.4 \cdot 1 + (-4.9) \cdot 5 + (-200.3) \cdot 150 = 0.4 - 24.5 - 30045 = -30069.1$$

Category 1:

$$\text{score}_1 = 0.3 \cdot 1 + 4.8 \cdot 5 + 200.5 \cdot 150 = 0.3 + 24 + 30075 = 30099.3$$

Category 2:

$$\text{score}_2 = (-0.1) \cdot 1 + 0.3 \cdot 5 + 0.2 \cdot 150 = -0.1 + 1.5 + 30 = 31.4$$

Step 3: Predict category

$$\hat{y}^{(3)} = \arg\max(-30069.1, 30099.3, 31.4) = 1$$

Step 4: Check if correct
True label: $y^{(3)} = 2$, Predicted: $\hat{y}^{(3)} = 1$
Prediction is WRONG! $\rightarrow$ Update weights
Step 5: Update weights
Increase row 2 (true category):

$$\mathbf{w}_2^{\text{new}} = [-0.1, 0.3, 0.2] + [1, 5, 150] = [0.9, 5.3, 150.2]$$

Decrease row 1 (predicted category):

$$\mathbf{w}_1^{\text{new}} = [0.3, 4.8, 200.5] - [1, 5, 150] = [-0.7, -0.2, 50.5]$$

Row 0 unchanged:

$$\mathbf{w}_0^{\text{new}} = [0.4, -4.9, -200.3]$$

Final weight matrix after all samples:

$$W = \begin{bmatrix} 0.4 & -4.9 & -200.3 \\ -0.7 & -0.2 & 50.5 \\ 0.9 & 5.3 & 150.2 \end{bmatrix}$$

(b) Predict for new user: Excitement = 6, Budget = 200

Step 1: Form input vector

$$\mathbf{x}^{\text{new}} = \begin{bmatrix} 1 \\ 6 \\ 200 \end{bmatrix}$$

Step 2: Calculate scores using final weights
Category 0:

$$\text{score}_0 = 0.4 \cdot 1 + (-4.9) \cdot 6 + (-200.3) \cdot 200 = 0.4 - 29.4 - 40060 = -40089$$

Category 1:

$$\text{score}_1 = (-0.7) \cdot 1 + (-0.2) \cdot 6 + 50.5 \cdot 200 = -0.7 - 1.2 + 10100 = 10098.1$$

Category 2:

$$\text{score}_2 = 0.9 \cdot 1 + 5.3 \cdot 6 + 150.2 \cdot 200 = 0.9 + 31.8 + 30040 = 30072.7$$

Step 3: Predict category

$$\hat{y}^{\text{new}} = \arg\max(-40089, 10098.1, 30072.7) = 2$$

Predicted Activity Category: 2 (Relaxation spots)