Name: ______________________________ **Roll Number:** ______________

~~40~~. Both Autoencoders and PCA can be used for dimensionality reduction. __(a)__ performs linear projections, whereas __(b)__ can learn nonlinear mappings through neural networks. While __(c)__ directly computes principal directions via eigenvalue decomposition, __(d)__ relies on gradient-based optimization and backpropagation. Unlike __(e)__, __(f)__ can include additional constraints such as sparsity or denoising objectives, making them more flexible but computationally more expensive. **[1 mark]**

i. a) Autoencoder, b) PCA, c) PCA, d) Autoencoder, e) PCA, f) Autoencoder
~~ii.~~ a) PCA, b) Autoencoder, c) PCA, d) Autoencoder, e) PCA, f) Autoencoder
iii. a) PCA, b) Autoencoder, c) PCA, d) Autoencoder, e) Autoencoder, f) PCA
iv. None of the others

41. Consider a feedforward neural network with two input neurons x_1 and x_2, a single hidden neuron, and one output neuron. The hidden neuron computes a pre-activation $a_h = w_1x_1 + w_2x_2 + b_1$ and uses a ReLU activation $z_1 = max(0, a_h)$. The output neuron computes a pre-activation $a_0 = w_3z_1 + b_2$ and applies a sigmoid activation $y = \sigma(a_0)$. Given the training sample $x_1 = 2$, $x_2 = 3$ with target $t = 1$, initial weights $w_1 = 0.5$, $w_2 =- 0.3$, $w_3 = 0.8$, biases $b_1 = b_2 = 0.2$, learning rate $\eta = 0.1$, ReLU at hidden, sigmoid at output, and loss $L = \frac{1}{2}(y - t)^2$, compute the updated weights after one gradient-descent update. For all numerical values, round to 3 decimal places. **[2 marks]**

i. $w_1 = 0.5$, $w_2 =- 0.3$, $w_3 = 0.8$
ii. $w_1 = 0.409$, $w_2 =- 0.285$, $w_3 = 0.802$
iii. $w_1 = 0.489$, $w_2 = 0.285$, $w_3 = 0.782$
iv. None of the others

X 42. Consider a fully connected feedforward neural network with:

- Input layer: 2 neurons
- One hidden layer: 2 neurons, activation function $\sigma(z) = \frac{1}{1+e^{-z}}$ (sigmoid)
- Output layer: 1 neuron, activation = sigmoid
- Loss: mean squared error $L = \frac{1}{2}(y - \hat{y})^2$

The parameters are:

$$W^{(1)} = \begin{bmatrix} 0.12 & -0.07 \\ 0.25 & 0.10 \end{bmatrix}; \; b^{(1)} = [0.05 \;\; - 0.02\,]$$

$$W^{(2)} = [0.40 \;\; - 0.30\,]; \; b^{(2)} = 0.10$$

A single training example is given as $x = [1,\ 2\,]$; $y = 0$. Use a learning rate $\eta = 0.2$.
Compute the network's output $\hat{y}$ for this input, then perform one step of gradient descent using backpropagation (i.e., compute gradients and update parameters). Report the following final results numerically (rounded to **4 decimal places**): **[3 marks]**

Name: ______________________________ **Roll Number:** ________________

a. $\hat{y}$ (network output before update)

b. Gradients: $\frac{\partial L}{\partial W^{(2)}}$, $\frac{\partial L}{\partial b^{(2)}}$, $\frac{\partial L}{\partial W^{(1)}}$, $\frac{\partial L}{\partial b^{(1)}}$

c. Updated parameters $W^{(1)}_{new}$, $b^{(1)}_{new}$, $W^{(2)}_{new}$, $b^{(2)}_{new}$

Treat rows of $W^{(1)}$ as the weights for each hidden neuron: $z^{(1)}_i = w^{(1)}_i \cdot x + b^{(1)}_i$.

43. You have a friend who only does one of four things on every Saturday afternoon: go shopping, watch a movie, play tennis, or just stay in. You have observed your friend's behavior over 11 different weekends. On each of these weekends, you have noted the weather (sunny, windy, or rainy), whether her parents visit (visit or no-visit), whether she has drawn cash from an ATM machine (rich or poor), and whether she had an exam during the coming week (exam or no-exam). You have built the following data table: **[3 marks]**

# ex.	Weather	Parents	Cash	Exam	Decision
1	sunny	visit	rich	yes	cinema
2	sunny	no-visit	rich	no	tennis
3	windy	visit	rich	no	cinema
4	rainy	visit	poor	yes	cinema
5	rainy	no-visit	rich	no	stay-in
6	rainy	visit	poor	no	cinema
7	windy	no-visit	poor	yes	cinema
8	windy	no-visit	rich	yes	shopping
9	rainy	visit	rich	no	shopping
10	sunny	no-visit	rich	no	tennis
11	sunny	no-visit	poor	yes	tennis
12	windy	visit	rich	yes	stay-in

You now want to build a decision tree to predict the activity of your friend on any future Saturday afternoon from the observed values of Weather, Parents, Cash, and Exam. Draw the full tree that correctly classifies all the examples.

~~44~~. Consider the following hypothetical data concerning student characteristics and whether or not each student should be hired. **[2 marks]**

Name	**GPA**	**Effort**	**Hirable**
Sarah	poor	lots	Yes
Dana	average	some	No
Alex	average	some	No
Annie	average	lots	Yes
Emily	excellent	lots	Yes
Pete	excellent	lots	No
John	excellent	lots	No
Kathy	poor	some	No

Use a Naive Bayes classifier to determine whether or not someone with excellent attendance, poor GPA, and lots of effort should be hired.

X 45. A dataset contains four 2-D samples: **[2 marks]**
$x_1 = [2\ 0]$; $x_2 = [0\ 2]$; $x_3 = [3\ 1]$; $x_4 = [1\ 3]$.

Perform PCA and report the following (round **all** final numeric answers to **4 decimal places**):

a. Mean vector μ and the centered data matrix.

b. Covariance matrix (2×2).

c. Eigenvalues and their corresponding normalized eigenvectors.

d. Which is the first principal component (direction) and the proportion of total variance it explains.