Homework 3 - ID3 Decision Tree Learning and PCA

Question 1 (30 points): ID3 Decision Tree Learning

You are provided with a training set of examples (see Figure 1). Which feature will you pick first to split the data as per the ID3 decision tree learning algorithm? Show all your work: compute the information gain for all four attributes and pick the best one.

Step 1: Entropy of the Entire Dataset

There are 14 total examples: 9 `Yes`, 5 `No`.

$$\begin{aligned} Entropy(S) &= -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) \\ &\approx -0.643 - 0.530 \\ &= 0.940 \end{aligned}$$

Step 2: Information Gain for Each Attribute

Outlook

- Sunny: 5 samples (2 Yes, 3 No)

$$Entropy = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

- Overcast: 4 samples (4 Yes)

$$Entropy = 0$$

- Rain: 5 samples (3 Yes, 2 No)

$$Entropy = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

$$\begin{aligned} Gain(S, \text{Outlook}) &= 0.940 - \left(\frac{5}{14}\cdot 0.971 + \frac{4}{14}\cdot 0 + \frac{5}{14}\cdot 0.971\right) \\ &= 0.940 - (0.347 + 0 + 0.347) = \mathbf{0.246} \end{aligned}$$

Temperature

- Hot: 4 samples (2 Yes, 2 No), Entropy = 1.000
- Mild: 6 samples (4 Yes, 2 No), Entropy = 0.918
- Cool: 4 samples (3 Yes, 1 No), Entropy = 0.811

$$\begin{aligned} Gain(S, \text{Temperature}) &= 0.940 - \left(\frac{4}{14} \cdot 1.000 + \frac{6}{14} \cdot 0.918 + \frac{4}{14} \cdot 0.811\right) \\ &= 0.940 - (0.286 + 0.393 + 0.232) = \mathbf{0.029} \end{aligned}$$

Humidity

- High: 7 samples (3 Yes, 4 No), Entropy = 0.985
- Normal: 7 samples (6 Yes, 1 No), Entropy = 0.591

$$\begin{aligned} Gain(S, \text{Humidity}) &= 0.940 - \left(\frac{7}{14} \cdot 0.985 + \frac{7}{14} \cdot 0.591\right) \\ &= 0.940 - (0.492 + 0.296) = \mathbf{0.152} \end{aligned}$$

Wind

- Weak: 8 samples (6 Yes, 2 No), Entropy = 0.811
- Strong: 6 samples (3 Yes, 3 No), Entropy = 1.000

$$\begin{aligned} Gain(S, \text{Wind}) &= 0.940 - \left(\frac{8}{14} \cdot 0.811 + \frac{6}{14} \cdot 1.000\right) \\ &= 0.940 - (0.463 + 0.429) = \mathbf{0.048} \end{aligned}$$

Final Decision

$$\boxed{\text{Best Attribute = Outlook (Information Gain = 0.246)}}$$

Question 2 (20 points): Principal Component Analysis

You are given three data points: $(-1, -1),\ (0, 0),\ (1, 1)$.

(a) First Principal Component (Without Eigen-Decomposition)

First, compute the mean:

$$\mu = \frac{1}{3}((-1, -1) + (0, 0) + (1, 1)) = (0, 0)$$

Since the data lies along the line $y = x$, the direction of maximum variance is:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(b) Projecting the Data onto the First Principal Component

Each point $\mathbf{x}$ is projected using: $\mathbf{v}_1^\top \mathbf{x}$

- $(-1, -1) \to \frac{1}{\sqrt{2}}(-1 + -1) = -\sqrt{2}$
- $(0, 0) \to \frac{1}{\sqrt{2}}(0 + 0) = 0$
- $(1, 1) \to \frac{1}{\sqrt{2}}(1 + 1) = \sqrt{2}$

1D Projected Data: $[-\sqrt{2},\ 0,\ \sqrt{2}]$

Question 3 (50 points): PCA Reconstruction Analysis Report

Results

Principal Components (p)	Reconstruction Error
10	155,351.4742
50	41,024.8648
100	14,285.8125
200	1,371.4145

Table 1: Reconstruction error with varying numbers of principal components

Key Findings

- **Error Reduction:**
  - Error decreases consistently as the number of components increases.
  - Most significant drop: **90.4%** from $p = 100$ to $p = 200$.
  - Initial drop: **73.6%** from $p = 10$ to $p = 50$.
- **Trade-off Between Quality and Compression:**
  - $p = 10$: Very high compression (96%), poor quality.
  - $p = 50$: Moderate compression (80%), acceptable quality.
  - $p = 100$: Balanced trade-off between compression (61%) and reconstruction quality.
  - $p = 200$: Near-lossless reconstruction (22% compression).

Conclusion

The PCA implementation demonstrates an effective trade-off between dimensionality reduction and reconstruction accuracy. While aggressive compression is possible with fewer components, maintaining high visual fidelity requires more principal components. For the given dataset, using around 100 components yields a good balance between compression and quality. Using 200 components achieves near-perfect reconstruction.

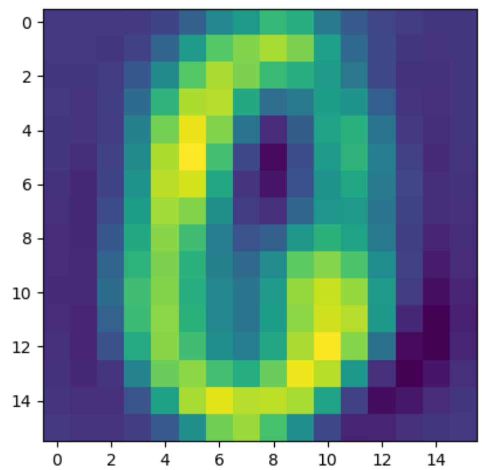

Figure 1: images1 with 10 component

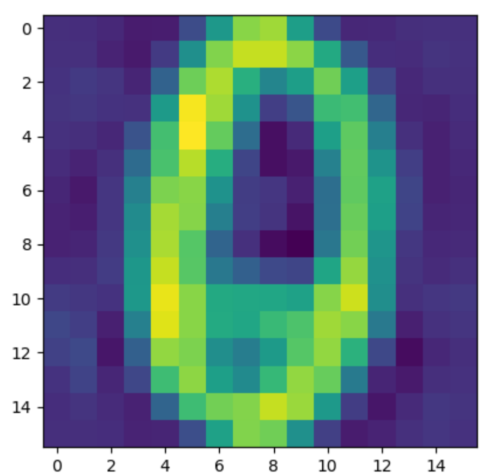

Figure 2: images1 with 50 component

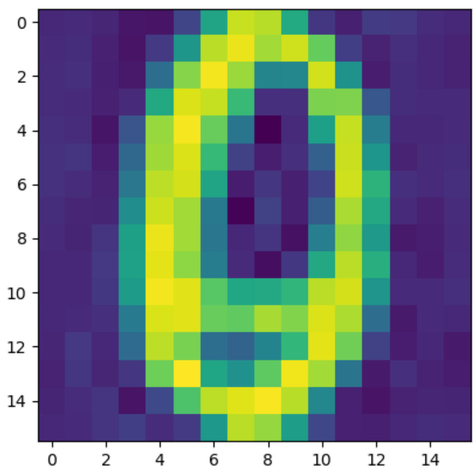

Figure 3: images1 with 100 component

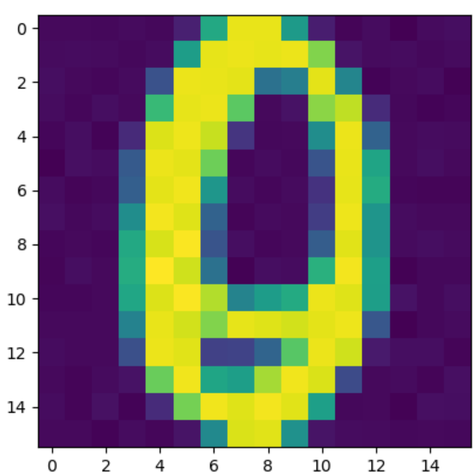

Figure 4: images1 with 200 component

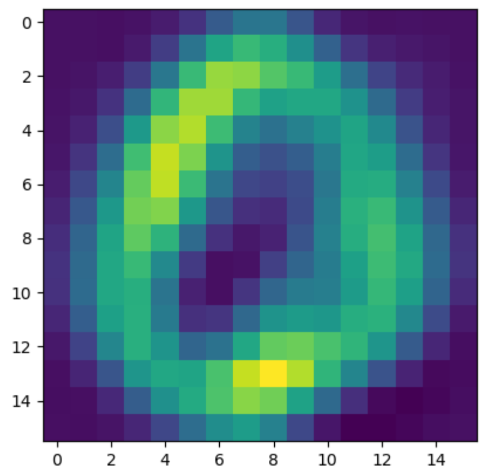

Figure 5: images2 with 10 component

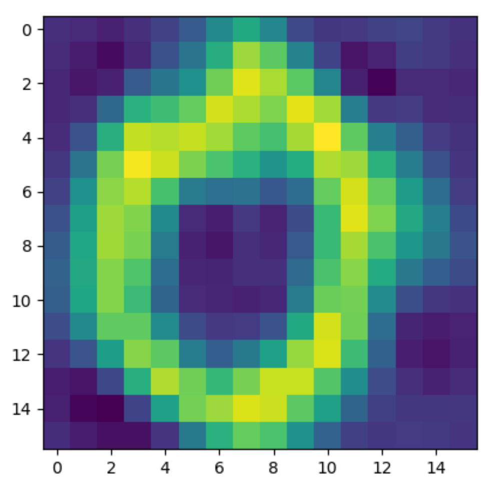

Figure 6: images2 with 50 component

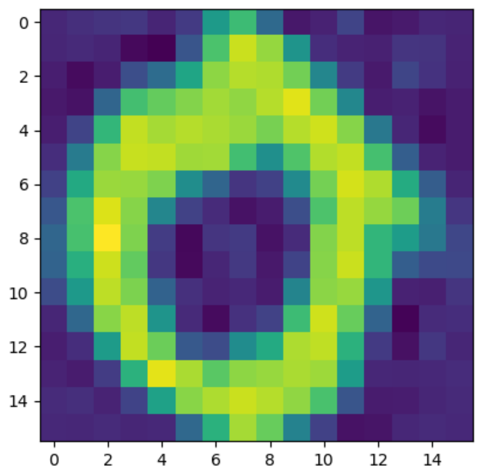

Figure 7: images2 with 100 component

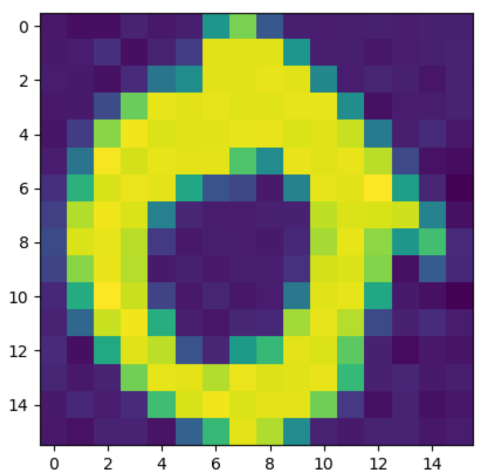

Figure 8: images2 with 200 component