Student Name: Weixin Liao
Student ID: 520021911014

Principle of Data Science

Assignment 1

---

1. Project Introduction

   The rapid growth of data in various domains, including computer vision and image processing, has created a demand for more efficient and accurate methods to process and analyze this data. One such domain is the classification of images based on their attributes. In this project, I aim to address the challenge of image classification using the Animals with Attributes 2 (AwA2) dataset, which consists of 37,322 images of 50 animal classes with pre-extracted deep learning features for each image.

   To achieve my objective, I employ a linear Support Vector Machine (SVM) classifier to classify images based on their deep learning features. However, the high dimensionality of these features can lead to increased computational complexity and potentially degrade the performance of the classifier. To mitigate this issue, I explore three different dimensionality reduction methods, including **feature selection** method, **feature projection** method, and **feature learning** method. I investigate their impact on the performance of the SVM classifier and identify the optimal dimensionality reduction method and the optimal dimensionality for the dataset.

   This report is structured as follows:

   (a) In Section 2, I present the methodology employed in my project, detailing the dataset, the training and testing split, the linear SVM classifier, and the dimensionality reduction methods.

   (b) Section 3 outlines my experimental results, presenting the performance of the SVM classifier with and without dimensionality reduction and comparing the different dimensionality reduction methods.

   (c) Section 4 concludes the report, summarizing the main outcomes of the project and suggesting potential future work.

2. Methodology

   In this section, I describe the methodology employed in my project, focusing on the AwA2 feature dataset, K-fold cross-validation for determining the best C parameter in the SVM classifier, and the various dimensionality reduction methods I have implemented.

   (a) Dataset

   The Animals with Attributes 2 (AwA2) dataset (https://cvml.ist.ac.at/AwA2/) is used for this project. It consists of 37,322 images of 50 animal classes, with pre-extracted deep learning features for each image. These features allow for a more efficient classification process compared to using raw image data.

   (b) Training and Testing Split

   To evaluate the performance of my classifier, I split the dataset into training and testing sets. I allocate 60% of the images in each class for training and the remaining 40% for testing. The K-fold cross-validation technique (with K=5) is employed within the training set to determine the optimal C parameter for the SVM classifier, which controls the trade-off between maximizing the margin and minimizing the classification error.

(c) Linear SVM Classifier

The linear Support Vector Machine (SVM) classifier is used for image classification based on the deep learning features. SVM is a powerful supervised learning algorithm that aims to find the optimal separating hyperplane between different classes in the feature space. By selecting the best C parameter through K-fold cross-validation, I ensure that the classifier achieves a balance between maximizing the margin and minimizing classification error.

(d) Dimensionality Reduction Methods

To reduce the computational complexity and improve the performance of the SVM classifier, I apply three different dimensionality reduction methods: feature selection method, feature projection method, and feature learning method. In each category, I implement two different algorithms, as detailed below:

i. **Sequential Forward Select** (SFG)

A algorithm that performs forward feature selection by iteratively adding features to the selected feature set. In each iteration, the algorithm evaluates the classification performance with the addition of each remaining feature, selecting the one that provides the best improvement. The process continues until the desired number of features is obtained. The Pseudo Code of SFG are as followed:

procedure SFG(X, y, k)
    $selected\_features \leftarrow \emptyset$
    while $|selected\_features| < k$ **do**
        $best\_performance \leftarrow -\infty$
        $best\_feature \leftarrow None$
        for $feature$ in X **do**
            if $feature \notin selected\_features$ **then**
                $temp\_features \leftarrow selected\_features \cup feature$
                $performance \leftarrow$ Evaluate classifier performance using $temp_features$
                if $performance > best\_performance$ **then**
                    $best\_performance \leftarrow performance$
                    $best\_feature \leftarrow feature$
                end if
            end if
        end for
        $selected\_features \leftarrow selected\_features \cup best\_feature$
    end while
    return $selected\_features$
end procedure

ii. **Recursive Feature Elimination** (RFE)

An iterative method that uses the SVM classifier's coefficients to rank and eliminate features. In each step, the least important features are removed, and the process is repeated until the desired number of features is obtained.

iii. Principal Component Analysis (PCA)

A linear dimensionality reduction technique that projects the data onto a lower-dimensional subspace by identifying the directions (principal components) that capture the most variance in the data.

The Pseudo Code of PCA are as followed:

procedure PCA(X, k)
    Compute the mean-centered data matrix X_{mc}
    Compute the covariance matrix $C = \frac{1}{n} X_{mc}^T X_{mc}$
    Perform eigenvalue decomposition of C: $C = Q \Lambda Q^T$

Select the first k eigenvectors corresponding to the largest eigenvalues: Q_k
Compute the lower-dimensional representation: $X_{PCA} = X_{mc}Q_k$
return X_{PCA}
end procedure

iv. **Autoencoder**
A type of neural network that learns to reconstruct the input data by encoding it into a lower-dimensional latent space and then decoding it back to the original space. The trained autoencoder's encoder part is used to obtain the lower-dimensional representation of the data.

v. **Sparse Coding** (SC)
A method that learns a dictionary of basis functions (atoms) and represents the data as sparse linear combinations of these atoms, using a sparse coding algorithm like LASSO LARS.

vi. **Locally Linear Embedding** (LLE)
A non-linear dimensionality reduction technique that aims to preserve the local structure of the data by approximating each data point as a linear combination of its nearest neighbors, and then embedding the data into a lower-dimensional space while maintaining these relationships.
The Pseudo Code of LLE are as followed:

procedure LLE($X, k, n_{neighbors}$)
  for i in $1, \ldots, n$ **do**
    Find the $n_{neighbors}$ nearest neighbors of x_i
    Compute the local covariance matrix C_i
    Compute the local weight matrix W_i by minimizing $||x_i - \sum_{j=1}^{n_{neighbors}} w_{ij}x_j||^2$ subject to $\sum_{j=1}^{n_{neighbors}} w_{ij} = 1$
  end for
  Compute the global matrix $M = (I - W)^T(I - W)$
  Perform eigenvalue decomposition of M: $M = Q\Lambda Q^T$
  Select the k eigenvectors corresponding to the smallest non-zero eigenvalues: Q_k
  Compute the lower-dimensional representation: $X_{LLE} = Q_k$
  return X_{LLE}
end procedure

3. Experiment Results

In this section, I will present the experimental results obtained from my tests. I applied various dimensionality reduction methods and tested their impact on the classification accuracy of the SVM classifier. The results are summarized in the table below:

From the results, we can observe that:

(a) When no dimensionality reduction is performed, the SVM classifier achieves the highest accuracy (0.93) on the original 2048-dimensional feature space.

(b) Sequential Forward Feature Selection (SFG) achieves comparable accuracy with the original feature space when the dimensionality is set to 512. At a lower dimensionality (128), the accuracy is slightly reduced to 0.90.

(c) Recursive Feature Elimination (RFE) performs similarly to SFG, with a slight improvement in accuracy at the lower dimensionality (128).

(d) Principal Component Analysis (PCA) and Autoencoder-based feature projection methods show consistent performance at different dimensionality levels (128 and 512) and maintain a high classification accuracy close to the original feature space.

Table 1: Experimental Results

Method	**Best_C**	**Dimensions**	**Accuracy**
None	1e-3	2048	0.93
SFG	1e-2	128	0.90
	1e-3	512	0.93
RFE	1e-2	128	0.91
	1e-3	512	0.93
PCA	1e-1	128	0.92
	1e-2	512	0.93
Autoencoder	1e-2	128	0.92
	1e-1	512	0.93
LLE	100	128	0.89
	200	512	0.91
Sparsecode	100	128	0.88

(e) Locally Linear Embedding (LLE) and Sparse Code (SC) exhibits lower classification accuracy compared to other methods, especially at lower dimensions (128).

Based on these observations, we can conclude that PCA and Autoencoder-based feature projection methods are the most robust and consistent dimensionality reduction techniques in my experiments. Both methods maintain high classification accuracy, comparable to the original feature space, regardless of the reduced dimensionality. This indicates that PCA and Autoencoder methods are effective at capturing the essential information in the data while significantly reducing the dimensionality.

On the other hand, SFG and RFE also perform well in terms of classification accuracy but show a slight reduction in performance when the dimensionality is lower (128). What's worse, the RFE runs so slowly in this project and it's unacceptable. This suggests that these feature selection methods might be more sensitive to the chosen dimensionality compared to PCA and Autoencoder methods. However, they still manage to achieve competitive results, making them viable options for dimensionality reduction depending on the specific application and computational constraints.

Lastly, LLE demonstrates the lowest classification accuracy among the tested methods, especially at the lower dimensionality (128) (and also runs too slowly). This suggests that LLE might not be the best choice for dimensionality reduction in this particular problem. It is worth noting that LLE is a non-linear dimensionality reduction technique, and its performance can vary greatly depending on the underlying structure of the data. In some cases, LLE might be more effective than linear methods like PCA, but in my experiments, it shows lower performance.

In conclusion, my experiments demonstrate that PCA and Autoencoder-based feature projection methods provide the best balance between dimensionality reduction and classification accuracy for the AwA2 feature dataset. These methods are effective at capturing the essential information while significantly reducing the dimensionality, making them suitable for a wide range of applications. Other methods, such as SFG and RFE, also show promising results, but their performance might be more sensitive to the chosen dimensionality.

4. Conclusion

In this report, I have presented a comprehensive study on the application of different dimensionality

reduction techniques for image classification using the Animals with Attributes 2 (AwA2) dataset and a linear Support Vector Machine (SVM) classifier. I investigated three primary categories of dimensionality reduction methods: feature selection, feature projection, and feature learning. Within each category, I evaluated two different algorithms.

The experimental results demonstrated the effectiveness of the dimensionality reduction methods in reducing the feature space while maintaining competitive classification accuracy. The best-performing methods were found to be Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and Autoencoder-based projection. All of them achieved an accuracy of 0.93 when applied to the dataset with reduced dimensions of 512. However, some methods, such as Sparse Coding and LLE, did not perform as well, suggesting that they may not be suitable for this particular dataset or task.

In conclusion, my study has shown that dimensionality reduction techniques can be an effective way to improve the computational efficiency of SVM classifiers without significant loss in performance. Moreover, the choice of the most suitable dimensionality reduction method should be made according to the specific dataset and problem. As the codes are also include on the assignment package, you can feel free to check any details you interested in on the codes file.

Potential future work could explore other dimensionality reduction techniques, such as t-SNE or UMAP, to assess their performance in similar tasks. Additionally, combining different dimensionality reduction methods or experimenting with ensemble techniques might lead to further improvements in classification accuracy. Lastly, investigating the impact of different SVM kernel functions and hyperparameters, as well as applying the proposed methodology to other classification algorithms, would provide valuable insights into the generalizability and effectiveness of the dimensionality reduction techniques in image classification tasks.