NYCU Introduction to Machine Learning, Homework 4

[111550086], [陳冠智]

Part. 1, Kaggle (70% [50% comes from the competition]):

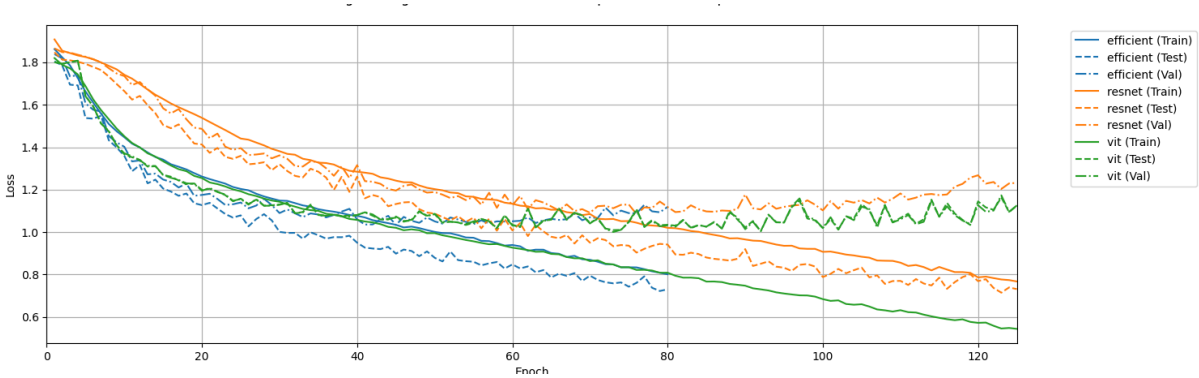
(10%) Implementation Details

Model backbone (e.g., VGG16, VGG19, Custom, etc)	1. EfficientNet-B0 (pretrained) 2. ResNet18 (from scratch) 3. Vision Transformer (ViT base patch16 pretrained)
Number of model parameters	1. EfficientNet-B0: ~5.3M 2. ResNet18: ~11.7M 3. ViT: ~86M
Other hyperparameters …	1.Epochs: 100 2. Learning rate: 1e-4 3. Weight decay: 0.01 4. Batch size: 64 5. Early stopping patience: 15 6. Data augmentation probability: 0.5

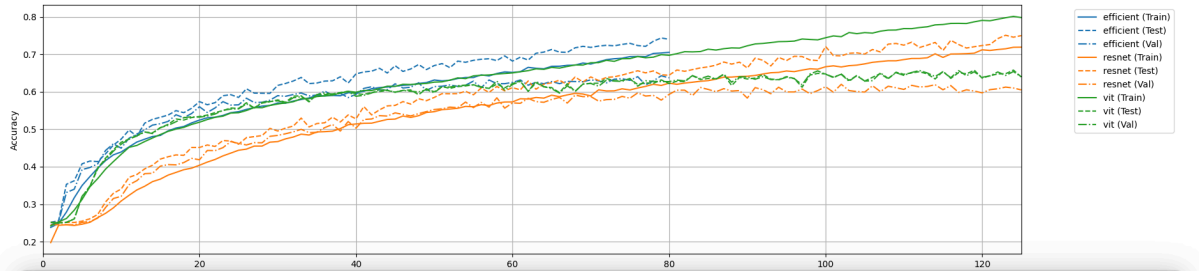
(10%) Experimental Results

Type your answer here.

(1) Loss



(2)accuracy



Part. 2, Questions (30%):

1. (10%) Explain the support vector in SVM and the slack variable in Soft-margin SVM. Please provide a precise and concise answer. (each in two sentences)

   Support vectors in SVM: The support vectors are the training points that lie closest to the decision boundary (hyperplane) and directly influence its position. They are critical points that define the margin of separation between classes, as removing them would change the position of the hyperplane.
   Slack variables in Soft-margin SVM: Slack variables allow some data points to violate the margin requirement or be misclassified, making SVM more practical for non-linearly separable real-world data. They represent the degree of violation from the margin constraint for each data point, with larger values indicating greater violation.

2. (10%) In training an SVM, how do the parameter C and the hyperparameters of the kernel function (e.g., γ for the RBF kernel) affect the model's performance? Please explain their roles and describe how to choose these parameters to achieve good performance.

Parameter C (Regularization Parameter):

- Controls the trade-off between having a large margin and minimizing classification errors
- Large C: Creates a smaller margin but fits training data more closely, risking overfitting
- Small C: Allows a larger margin but may increase training errors, potentially underfitting
- Think of C as controlling how much we penalize misclassifications during training

Kernel Parameter γ (RBF Kernel):

- Controls the influence radius of each support vector
- Large γ : Creates a smaller influence radius, leading to more complex decision boundaries
- Small γ : Creates a larger influence radius, resulting in smoother decision boundaries
- Think of γ as controlling how "local" or "global" the model's view of the data is

How to choose these parameters:

1. Use Grid Search or Random Search with cross-validation to find optimal values
2. Start with a logarithmic scale (e.g., C = 0.1, 1, 10, 100; γ = 0.01, 0.1, 1)
3. Monitor both training and validation performance to avoid overfitting
4. Once you find a good region, perform a finer search around those values

A practical approach is to:

1. First try a coarse grid search to identify promising regions
2. Then perform a finer search in those regions
3. Always validate the model's performance on a separate validation set
4. Consider the problem's characteristics (e.g., noise level, data complexity) when setting search ranges

.

3. (10%) SVM is often more accurate than Logistic Regression. Please compare SVM and Logistic Regression in handling outliers.

SVM:

- SVM is more robust to outliers because it only depends on support vectors (points near the decision boundary), not all data points
- In soft-margin SVM, the parameter C controls how much influence outliers have on the decision boundary, allowing us to reduce their impact even further

Logistic Regression:

- Logistic Regression is more sensitive to outliers because it uses all data points to find the decision boundary
- It tries to minimize the total error across all points, which means outliers can significantly pull or distort the decision boundary

Key Differences in Outlier Handling:

1. Influence Range:
    - SVM: Only affected by support vectors
    - Logistic Regression: Affected by all points
2. Impact Level:
    - SVM: Outliers have limited impact due to margin-based optimization
    - Logistic Regression: Outliers can have large impact due to the global nature of its optimization

This difference in outlier handling is one of the main reasons why SVM often achieves better accuracy than Logistic Regression in real-world datasets where outliers are common.