Homework 3 Report: Logistic Regression with Regularization

1. Solution

1.1 Logistic Regression with Regularization
Logistic regression is a supervised learning algorithm used for binary classification. It models the probability that a given input belongs to the positive class by computing a weighted sum of the input features and passing this value through a logistic function to produce a probability between 0 and 1.

During training, the model minimizes a cross-entropy loss that measures the difference between predicted probabilities and true class labels. To prevent overfitting, an L2 regularization term is added to the loss function, penalizing large weight values. This penalty improves stability, smooths the decision boundary, and reduces the influence of noisy features. The weights are iteratively updated through gradient-based optimization until convergence. After training, predicted probabilities greater than 0.5 are classified as the positive class, while lower values are assigned to the negative class.

1.2 Data Transformation
Before training, input features were standardized so that each feature has a mean of 0 and a standard deviation of 1. Standardization ensures all variables contribute proportionally to the model and improves numerical stability during optimization.

To allow the model to capture nonlinear relationships, the feature space was expanded using polynomial transformations of degrees 1, 2, and 3.

- Degree 1: Original standardized features.
- Degree 2: Adds squared and pairwise interaction terms.
- Degree 3: Adds cubic and higher-order interactions.

Each expanded feature set defines a separate hypothesis space. These were all evaluated under the same training and validation process to compare how feature complexity affects generalization performance.

1.3 Legendre Polynomials for Feature Transform
As an alternative to ordinary polynomial expansion, Legendre polynomials were used to generate orthogonal feature transformations. Each feature was first scaled to the range [–1, 1] and then

mapped to Legendre bases up to degree 3. The orthogonality property of Legendre polynomials reduces redundancy between higher-order terms and improves the numerical conditioning of the model. Legendre polynomials were applied independently to each feature without generating cross-feature interaction terms.

This transformation makes the effect of regularization more consistent and predictable while retaining the ability to model nonlinear relationships. The Legendre-based transformations were evaluated using the same cross-validation framework as the standard polynomial features.

---

2. 5-Fold Cross Validation

2.1 Experiment Setup

The experiment was conducted using the Breast Cancer Wisconsin dataset, which contains 569 samples with 30 numerical features and two class labels. Each feature was standardized prior to model training. Logistic regression was applied with L2 regularization to examine the effect of different regularization strengths and feature transformations.

The regularization parameter λ was tested across multiple values (0.01, 0.1, 1, 10, and 100) to observe how varying the penalty influences accuracy and stability. For each λ, a 5-fold cross-validation procedure was used to estimate validation performance.

In 5-fold cross validation, the dataset is divided into five equal partitions. In each iteration, four folds are used for training and one fold for validation. The process repeats five times so that each fold serves once as the validation set. The average accuracy across the five runs provides a reliable estimate of model performance and helps reduce variability caused by random data splits.

Feature transformations of polynomial degree 1, 2, and 3 were each evaluated separately under identical conditions. The same setup was also applied to the Legendre polynomial transformations to compare their effect on stability and generalization.

2.2 Results

The dataset contains 569 samples and 30 numerical features representing diagnostic measurements. The class distribution consists of 212 samples labeled as class 0 and 357 samples labeled as class 1.

For standard polynomial transformations:

- Accuracy generally improved when moving from degree 1 to degree 2, indicating that the additional nonlinear terms helped capture meaningful feature interactions.
- Degree 3 provided only marginal improvement and, in some cases, showed minor instability when regularization was weak (small λ values).
- The best overall performance occurred at moderate regularization levels, typically $\lambda = 0.1$ or $\lambda = 1$, which balanced bias and variance effectively.

For the Legendre polynomial transformations:

- Accuracy trends were similar to those of standard polynomial features but displayed smoother curves across different λ values.
- The orthogonality of the Legendre basis reduced variability and made performance less sensitive to changes in regularization.
- Overall, Legendre transformations achieved comparable accuracy with slightly improved stability.

The accuracy vs. λ plots confirmed these patterns, showing that regularization strength plays a crucial role in controlling model complexity and ensuring consistent performance across feature transformations.

Dataset Summary:

X shape: (569, 30)
y shape: (569,)
Classes: [0 1]
Original shape: (569, 30)
Polynomial (2) shape: (569, 495)

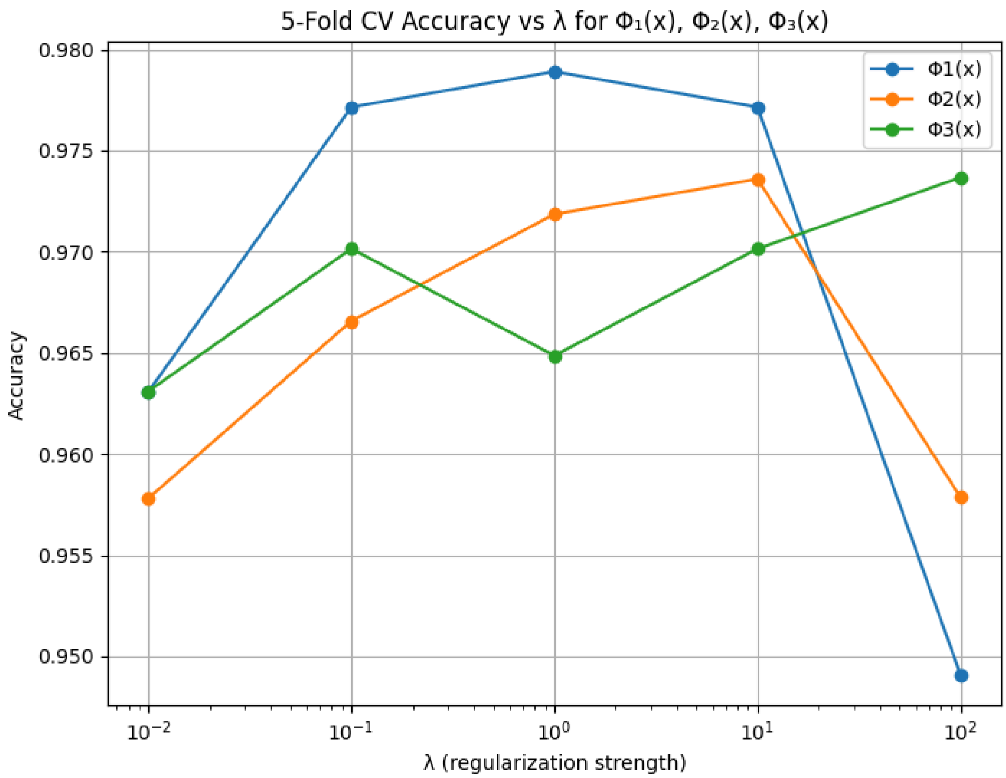
Polynomial Feature Transformations:

$\Phi1(x)$: Polynomial degree 1
$\lambda = 0.01$ → Accuracy = 0.9631
$\lambda = 0.1$ → Accuracy = 0.9771
$\lambda = 1$ → Accuracy = 0.9789
$\lambda = 10$ → Accuracy = 0.9771
$\lambda = 100$ → Accuracy = 0.9491

$\Phi2(x)$: Polynomial degree 2
$\lambda = 0.01$ → Accuracy = 0.9578
$\lambda = 0.1$ → Accuracy = 0.9666
$\lambda = 1$ → Accuracy = 0.9719
$\lambda = 10$ → Accuracy = 0.9736
$\lambda = 100$ → Accuracy = 0.9578

$\Phi3(x)$: Polynomial degree 3
$\lambda = 0.01 \rightarrow$ Accuracy = 0.9631
$\lambda = 0.1 \rightarrow$ Accuracy = 0.9701
$\lambda = 1 \rightarrow$ Accuracy = 0.9649
$\lambda = 10 \rightarrow$ Accuracy = 0.9701
$\lambda = 100 \rightarrow$ Accuracy = 0.9737

Best Configuration:
Best λ: 10
Best Accuracy: 0.9736

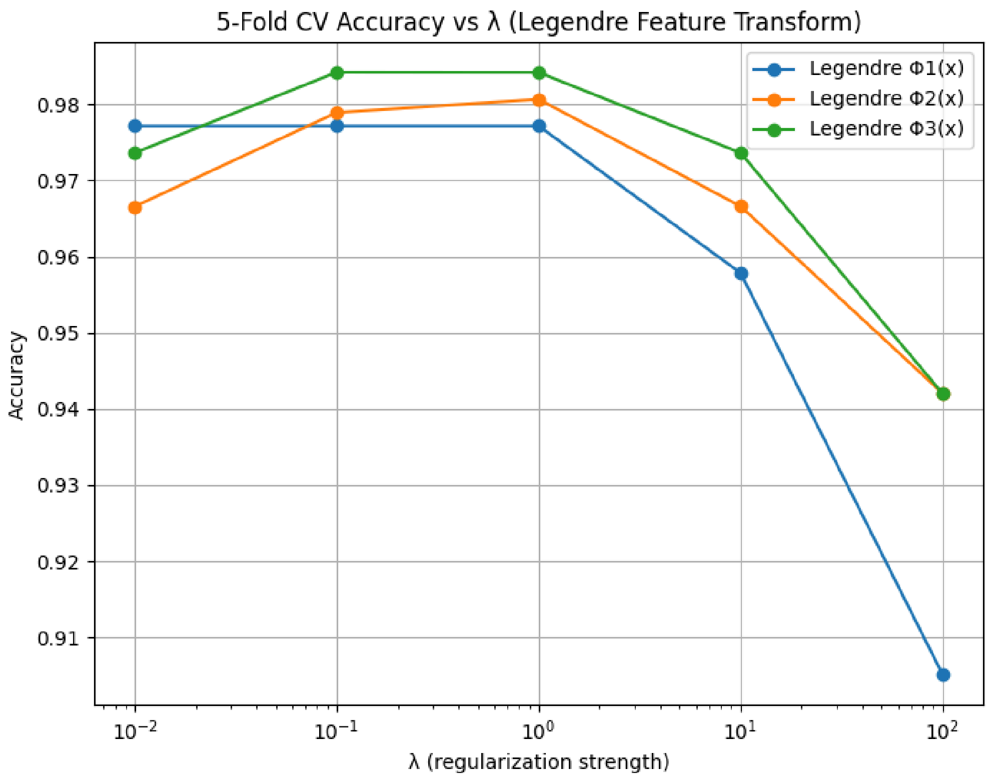


Legendre Polynomial Feature Transformations:
Legendre $\Phi1(x)$: degree 1
$\lambda = 0.01 \rightarrow$ Accuracy = 0.9771
$\lambda = 0.1 \rightarrow$ Accuracy = 0.9771
$\lambda = 1 \rightarrow$ Accuracy = 0.9771
$\lambda = 10 \rightarrow$ Accuracy = 0.9578
$\lambda = 100 \rightarrow$ Accuracy = 0.9051

Legendre $\Phi2(x)$: degree 2
$\lambda = 0.01 \rightarrow$ Accuracy = 0.9666
$\lambda = 0.1 \rightarrow$ Accuracy = 0.9789

λ = 1 → Accuracy = 0.9807
λ = 10 → Accuracy = 0.9666
λ = 100 → Accuracy = 0.9420

Legendre Φ3(x): degree 3
λ = 0.01 → Accuracy = 0.9736
λ = 0.1 → Accuracy = 0.9842
λ = 1 → Accuracy = 0.9842
λ = 10 → Accuracy = 0.9736
λ = 100 → Accuracy = 0.9420



2.3 Discussion

The cross-validation results demonstrate that L2 regularization successfully prevents overfitting in logistic regression, particularly when higher-degree feature transformations are introduced. Models without sufficient regularization tend to overfit the training data, while those with excessive regularization underfit and lose accuracy.

Polynomial transformations expanded the model's flexibility and improved accuracy by capturing nonlinear relationships. However, higher-degree expansions also increased the risk of overfitting, emphasizing the need for balanced regularization.

The Legendre transformation provided a more stable alternative by introducing orthogonal features that reduced correlation among high-order terms. This improved numerical conditioning and made performance less dependent on fine-tuning λ.

Overall, the experiments confirmed that moderate regularization with second-degree transformations achieved the best trade-off between accuracy, complexity, and stability.