

ML for Smart Agriculture

Ridge Regression and Logistic Regression

Due: Friday, 2nd August,
Late Deadline: Monday 5th August

1. Ridge Regression

- (a) Download PS2-1 dataset. Write code in Python that applies Ridge regression to the dataset to compute θ for given λ . Implement two cases i) closed-form solution and ii) stochastic gradient descent with mini-batch of size m . [20 Points]
- (b) Implement K-fold cross validation on the training set to obtain best regularization λ and get optimal θ . Consider root mean squared error (RMSE) as regression error, and report error on test samples. Report optimal λ , θ test and training errors for $K \in \{2, 10, N\}$, where N is number of samples. For all cases, consider n -degree polynomials, and basis function expansion $\phi(\cdot) = [1, x, x^2, \dots, x^n]$, try $n \in \{2, 5, 10\}$. [10 Points]

2. Logistic Regression

- (a) Write code in python that takes input a training dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, and its output is the weight vector \mathbf{w} in the logistic regression model $y = \sigma(\mathbf{w}^T \mathbf{x})$. [15 Points]
- (b) Download the dataset PS2-2. Run (a) on training data to compute \mathbf{w} and evaluate on test set. Plot the data (use different colors for data for different classes) and plot the decision boundary. [5 Points]
- (c) Repeat (b) using PS2-3. Explain the differences between two datasets and justify your results / observations. [5 Points]