Exercises on k-NN Classification

Exercise 1: k-NN with Euclidean Distance and Cross-Validation

Goal: Implement k-NN using Euclidean distance and evaluate performance using 5-fold cross-validation.

Dataset:

SepalLength	SepalWidth	PetalLength	PetalWidth	Class
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
5.8	4.0	1.2	0.2	setosa
6.0	2.2	4.0	1.5	versicolor
6.1	2.8	4.7	1.4	versicolor
5.9	3.0	4.2	1.5	versicolor
6.7	3.1	4.4	1.4	versicolor
6.3	2.5	4.9	1.5	versicolor
6.5	3.0	5.1	2.0	virginica
6.2	2.8	4.5	1.5	versicolor
6.4	2.9	4.3	1.3	versicolor
5.5	2.4	4.0	1.3	versicolor
5.7	2.8	4.1	1.3	versicolor
5.8	2.7	5.1	1.9	virginica
6.9	3.1	5.4	2.3	virginica
6.0	2.2	5.0	1.5	virginica
6.3	2.3	5.6	2.4	virginica
6.1	2.8	5.6	2.4	virginica
5.6	2.9	3.6	1.3	versicolor
5.8	2.7	4.1	1.0	versicolor
6.0	2.9	4.5	1.5	versicolor
6.1	2.6	4.7	1.4	versicolor
6.5	3.0	5.2	2.0	virginica
6.2	2.9	5.4	2.3	virginica
5.9	3.0	5.1	1.8	virginica
6.3	2.7	5.6	2.1	virginica

Tasks

Step 1: Data Preparation

(a) Take input of the data.

(b) Split into training (80%) and testing (20%) with a fixed random seed.

Step 2: Euclidean Distance

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Step 3: Cross-Validation

(a) Implement 5-fold cross-validation.

(b) For odd values of $k = 1, 3, 5, 7, 9$, compute the mean accuracy across folds.

(c) Print results in a formatted table (with 2 decimal places).

Exercise 2: k-NN with Manhattan Distance, K-Fold Validation, and Normalization

Dataset:

SepalLength	SepalWidth	PetalLength	PetalWidth	Class
5.2	3.4	1.5	0.2	setosa
4.8	3.1	1.6	0.3	setosa
5.0	3.2	1.2	0.2	setosa
5.3	3.7	1.4	0.3	setosa
4.9	3.0	1.5	0.1	setosa
5.1	3.5	1.3	0.3	setosa
5.4	3.4	1.7	0.2	setosa
5.0	3.3	1.4	0.2	setosa
6.0	2.7	4.2	1.3	versicolor
6.2	2.9	4.3	1.3	versicolor
5.7	2.6	3.5	1.0	versicolor
5.8	2.7	4.1	1.2	versicolor
6.1	3.0	4.6	1.4	versicolor
5.6	2.8	4.0	1.3	versicolor
6.3	2.5	4.9	1.5	versicolor
6.0	3.4	4.5	1.6	versicolor
5.9	3.0	4.2	1.5	versicolor
6.4	2.8	5.0	1.7	versicolor
5.5	2.5	4.0	1.2	versicolor
6.2	2.2	4.8	1.8	versicolor
6.5	3.0	5.2	2.0	virginica
6.9	3.1	5.4	2.1	virginica
6.7	3.0	5.8	2.2	virginica
7.1	3.0	5.9	2.1	virginica
6.3	2.9	5.6	1.8	virginica
6.6	2.8	5.3	2.0	virginica
7.0	3.2	5.7	2.3	virginica
6.5	3.2	5.1	2.0	virginica
6.8	3.0	5.5	2.1	virginica
6.4	2.9	5.6	2.2	virginica
6.2	3.4	5.4	2.3	virginica
6.9	3.1	5.1	2.3	virginica
7.2	3.2	6.0	2.2	virginica
6.3	2.8	5.7	1.9	virginica
6.1	3.0	5.5	1.8	virginica
6.7	3.3	5.7	2.1	virginica
6.4	3.1	5.5	1.8	virginica
6.8	3.2	5.9	2.3	virginica
7.3	2.9	6.1	2.5	virginica
6.5	3.0	5.8	2.2	virginica

Goal: Implement k-NN with Manhattan distance, normalize features, and evaluate performance using k-fold validation.

Tasks

Step 1: Data Normalization

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Step 2: Manhattan Distance

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

Step 3: K-Fold Validation

(a) Implement 5-fold cross-validation.

(b) For odd values of $k = 1, 3, 5, 7, 9$, compute the mean accuracy.

(c) Print results in a formatted table (with 2 decimal places).