
MA722 - Foundations of Machine Learning Algorithms

Lab examination

1. A food-processing company wants to understand how the internal temperature of a storage unit influences the spoilage score of a perishable product. The spoilage score is measured using a chemical degradation test. Fit a simple linear regression model to quantify how temperature affects spoilage. Additionally, examine model residuals to comment on linearity and noise.

Dataset

Temperature (°C)	Spoilage Score
12.5	22.4
13.2	23.5
14.1	25.0
15.4	25.7
16.0	27.2
17.5	28.1
18.3	30.0
19.2	31.4

2. A real-estate analytics firm wants to build a predictive model for the selling price of apartments based on structural features. Fit a multiple linear regression using square footage, number of bedrooms, distance to nearest metro, and age of property as predictors.

Dataset

Sqft	Bedrooms	Distance(km)	Age(yrs)	Price(lakhs)
900	2	0.8	5	48
1050	2	1.2	3	56
1120	3	0.5	8	62
1300	3	0.9	10	70
1450	3	0.6	6	78
1600	4	1.3	4	90

3. A drone tracks altitude versus control-thrust values. Noise in readings motivates using an iterative gradient-based least squares solver. Implement gradient descent and compare iterations with normal-equation solution.

Dataset

Thrust(N)	Altitude(m)
15	32
18	40
20	46
23	51
25	59
28	66

4. A subscription platform tracks whether a user renews based on time spent weekly, number of interactions, and customer-tier. Fit logistic regression and interpret coefficients.

Dataset

Minutes	Interactions	Tier(1-3)	Renew(0/1)
45	12	1	0
80	20	2	1
95	25	2	1
120	32	3	1
60	10	1	0
110	30	3	1

5. A bioinformatics lab classifies proteins into two structural classes using two geometric descriptors. Train an RBF SVM, perform grid-search over C and gamma, and visualize validation accuracy heatmap.

Dataset

FeatureA	FeatureB	Class
1.2	2.4	0
2.1	1.9	0
2.8	3.0	1
3.5	3.2	1
4.0	3.8	1
1.0	1.5	0

6. A manufacturing unit monitors eight correlated sensor readings. Apply PCA to reduce dimension to two principal components and interpret variance explained.

Dataset

S1	S2	S3	S4	S5	S6	S7	S8
3.2	4.1	2.5	5.2	3.8	4.0	3.1	4.4
3.5	4.4	2.7	5.0	4.0	4.2	3.0	4.7
2.9	3.8	2.2	4.8	3.5	3.9	2.8	4.0

7. A bank classifies loan applications as approved or rejected based on applicant features. Fit a decision tree, plot it, and compare entropy vs. Gini criteria.

Dataset

Income	Age	Score	Approved
45	30	700	1
55	40	650	1
35	28	600	0
60	45	720	1
30	25	580	0

8. A machine tool's failure time follows an exponential distribution. Using a prior on the rate parameter, compute the MAP estimate and compare with MLE.

Dataset

FailureTime(hrs)
21
18
23
25
19

9. A clinic evaluates whether symptoms (fever, cough) and test result influence probability of a respiratory infection. Construct a Bayesian network and compute posterior probabilities.

Dataset

Fever	Cough	TestPositive
1	1	1
0	1	0
1	0	1
0	0	0

10. A transport authority clusters city intersections based on traffic volume and signal wait times. Run K-means for $k = 2, 3, 4$ and compare silhouette scores.

Dataset

TrafficVol	WaitTime(s)
120	45
150	50
400	90
420	95
130	48
380	85
