
MA722 - Foundations of Machine Learning Algorithms

Lab examination

1. A pharmaceutical company studies how the concentration of an active compound affects the reaction time of a chemical assay. The goal is to fit a simple linear regression model to determine whether reaction time increases proportionally with concentration. Also construct a 95% confidence interval for the slope.

Dataset

Concentration(mg/mL)	ReactionTime(s)
2.0	5.2
2.5	6.0
3.0	6.8
3.5	7.5
4.0	8.4
4.5	9.1
5.0	9.8

2. A telecom provider wants to predict customer monthly bill based on call minutes, data usage, roaming usage, and customer tenure. Fit a multiple regression model, evaluate VIF for multicollinearity, and calculate adjusted R^2 .

Dataset

CallMin	DataGB	RoamMin	Tenure(yrs)	Bill(Rs)
300	2.5	10	1	580
450	4.0	20	2	750
500	5.5	15	3	820
600	6.2	25	4	910
700	7.0	30	5	1020

3. (a) Solve linear regression for small dataset via closed form and via gradient descent; compute ℓ_2 norm difference.
(b) Plot cost vs iterations for two learning rates; comment on convergence.
(c) Implement stochastic gradient descent; compare time-to-converge and final loss with batch GD.

Dataset (Synthetic - Regression Convergence)

Y	X1	X2	X3	X4	X5	NoiseVar (std)
10.2	1.0	2.5	0.5	3.0	1.2	0.8
15.4	2.1	3.1	1.0	4.2	0.9	1.0
8.6	0.6	1.8	0.2	2.1	1.5	0.7
20.3	3.0	4.0	1.5	5.3	0.8	1.2
12.0	1.4	2.9	0.9	3.4	1.1	0.9
7.8	0.5	1.2	0.1	1.9	1.8	0.6
18.7	2.8	3.7	1.3	4.9	0.7	1.1
9.5	0.9	2.0	0.4	2.7	1.4	0.75

4. A warehouse monitors whether a package is delivered on time (0/1) depending on distance, weight, and weather delay index. Fit logistic regression and compute confusion matrix.

Dataset

Distance(km)	Weight(kg)	DelayIndex	OnTime(0/1)
5	2	0.1	1
12	4	0.3	1
20	6	0.5	0
25	5	0.4	0
8	3	0.2	1

5. A meteorology lab classifies weather events as safe or high-risk based on humidity and wind speed. Fit an SVM with RBF kernel, perform grid-search over C and gamma, and plot the resulting validation heatmap.

Dataset

Humidity(%)	Wind(km/h)	Risk(0/1)
40	10	0
55	15	0
65	25	1
70	30	1
80	35	1
45	12	0

6. A transport fleet logs engine performance using six correlated variables. Apply PCA to reduce dimensionality to two components and evaluate the variance explained.

Dataset

E1	E2	E3	E4	E5	E6
3.0	4.2	5.1	6.0	4.8	5.5
3.3	4.0	4.9	5.8	5.0	5.2
2.9	4.1	5.0	5.9	4.7	5.3

7. A university predicts whether a student passes a course based on attendance, assignment scores, and midterm marks. Fit a decision tree, compare depth levels, and interpret the rules.

Dataset

Attend(%)	AssignScore	Midterm	Pass(0/1)
70	14	22	0
80	15	25	1
90	18	30	1
60	12	20	0
85	16	28	1

8. A call center models time between calls as an exponential distribution with a Gamma prior on the rate parameter. Compute the MAP estimate of the rate λ and compare with the MLE.

Dataset

InterArrival(min)
3.2
2.8
3.6
4.0
3.4

9. A medical diagnostic system evaluates three binary variables: presence of fatigue, blood-test abnormality, and disease presence. Construct the conditional probability table for a Bayesian network and compute posterior probability of disease.

Dataset

Fatigue	BloodAbnormal	Disease
1	1	1
1	0	0
0	1	1
0	0	0

10. A fitness club segments customers based on weekly workout time and calories burnt per session. Perform K-means for $k = 2, 3, 4$ and compare inertia values to determine the appropriate cluster count.

Dataset

WorkoutHrs	Calories
3.5	320
4.0	350
7.0	550
6.5	530
3.8	340
7.2	580
