Department of Mathematical and Computational Sciences
National Institute of Technology Karnataka, Surathkal

https://sam.nitk.ac.in/ sam@nitk.edu.in

MA722 - Foundations of Machine Learning Algorithms

Lab problem sheet 5

1. An automotive company wants to predict a car's fuel efficiency (Miles Per Gallon, MPG) based on its engine size. The relationship between engine size and MPG is nonlinear — larger engines tend to lower fuel efficiency, but not in a perfectly straight line. Fit a polynomial regression model that best captures this relationship and use it to make predictions.

   Dataset:

Engine_Size	MPG
1.0	42
1.3	39
1.5	35
1.8	31
2.0	29
2.3	26
2.5	24
3.0	20
3.5	18
4.0	15

   Tasks:

   (a) Load the dataset using `pandas`.
   (b) Visualize the data (scatter plot of Engine_Size vs MPG).
   (c) Fit a Linear Regression model and plot the regression line.
   (d) Fit a Polynomial Regression model (degree 3 or higher) and plot the curve on the same graph for comparison.
   (e) Predict the MPG for a car with an engine size of 2.2 using both models.
   (f) Evaluate and print the R^2 score for both models — which one fits better?

2. A coin is tossed 10 times and 7 heads are observed. Assume the coin's probability of landing heads, θ, follows a $\mathrm{Beta}(1, 1)$ prior (a uniform prior).

   Task:

   (a) Compute the posterior distribution
   $$P(\theta \mid \text{data}) \propto \theta^h (1-\theta)^t$$
   manually.
   (b) Normalize it so that the area under the posterior equals 1.

(c) Plot the prior, likelihood, and posterior on the same graph.

3. Use the *Iris* dataset to classify flower species using Gaussian Naive Bayes.

   Task:

   (a) Load the dataset using `sklearn.datasets.load_iris()`.
   (b) Split it into training and testing sets.
   (c) Train a `GaussianNB` model from `sklearn.naive_bayes`.
   (d) Predict on the test set and print the accuracy score.

4. Consider a Bayesian Network with three binary variables:

   - R: It is raining
   - S: The sprinkler is on
   - W: The grass is wet

   The dependencies are:

$$P(R) = 0.2$$

$$P(S \mid R) = 0.01, \quad P(S \mid \neg R) = 0.4$$

$$P(W \mid S, R) = 0.99, \quad P(W \mid S, \neg R) = 0.9, \quad P(W \mid \neg S, R) = 0.8, \quad P(W \mid \neg S, \neg R) = 0.0$$

   Tasks:

   (a) Compute the joint probability $P(R, S, W)$ for all combinations of truth values.
   (b) Compute the marginal probability $P(W)$.
   (c) Compute the conditional probability $P(R \mid W)$ using Bayes' rule.

5. For modeling a medical diagnosis scenario:

   - D: Has Disease
   - T: Test Result (Positive/Negative)

   The dependencies are:

$$P(D) = 0.01$$

$$P(T = + \mid D) = 0.95, \quad P(T = + \mid \neg D) = 0.1$$

   Tasks:

   (a) Create a Bayesian model with these dependencies.
   (b) Define the Conditional probability distriutions..
   (c) Compute:
      - $P(D)$
      - $P(D \mid T = +)$
   (d) Print and interpret the results.

*****