Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-II

Date: 11-09-2015

1. (a) Analyze:
   i. Data1 using linear regression.
   ii. Data2 and Data3 using linear regression & ridge regression.

   (b) Report the pre-processing and validation techniques used.

   (c) Plot $J(w)$ against iteration for all data and $J(w)$ against w for Data 3.

   (d) Report the parameters of the model.

   (e) Assess the performance of the model.

   (f) Plot the hyperplane that generates the data.

   (g) Compare the performance of linear and ridge regression.

2. (a) Develop 3rd, 6th and 7th degree polynomial models for polynomialdata.

   (b) Report the pre-processing and validation techniques used.

   (c) Plot λ (regularization parameters values) against training error and validation error, in a single figure. Plot J(w).

   (d) Report the parameters of the model.

   (e) Plot the graph of the resulting models.

   (f) Compare the performance of the models and select the best among them.

   [**Data1, Data2, Data3 and polynomial data are attached in the folder.**]

3. Analyze Boston Housing Data (download from UCI web repository)

   (a) Discuss the experimental results.

4. Solve the following linear equations using Gauss elimination

$$\begin{aligned} x+y-z &= 9 \\ 8y+6z &= -6 \\ -2x+4y-6z &= 40 \end{aligned}$$

   (a) Represent the equation in matrix form: $Ax = y$. Does y lies in the range space of A? Find the dimension and basis of range space of A.

5. Find the rank, basis and dimension of the range space of the following matrix:

$$X = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 2 & 4 & 2 \\ 0 & 2 & 2 & 1 \end{bmatrix}$$

6. Explain the term: subspace of a vector space, span of a set of vectors. Check whether:

   (a) Range space of matrix transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ a subspace of $\mathbb{R}^m$.

   (b) All vectors in $\mathbb{R}^3$ with $v_1 - v_2 + 2v_3 = 0$ a subspace of $\mathbb{R}^3$.

   (c) All vectors in $\mathbb{R}^2$ with $v_1 \geq v_2$ a subspace of $\mathbb{R}^2$. Here $[v_1, v_2, v_3]$ are components of a vector.

7. Check whether the set of vectors [0 1 1], [1 1 1], [0 0 1] are linearly independent.

8. Decribe:

   - Probability distribution, Multivariate Normal distribution, Unbiased Estimator. Find the unbiased estimator for the parameter of Bernoulli distribution.

Notes

- All the files related with the assignment should be saved in a single folder and send to sumitra@iist.ac.in.
- Submission dates: First phase submission: 17-09-2015; Second phase submission: 22-09-2015.
- **As far as assignments are concerned, students are expected to observe academic honesty and integrity. Though the students can collaborate and**

discuss, copying directly other students' assignment or allowing your own assignment to be copied constitute academic dishonesty and is highly discouraged.