Midterm Exam #2

MAT325 Numerical Analysis

2023-04-14

Three problems in the exam are related to the direct methods of linear algebra and their applications in the least square approximation of a function with either known or unknown expression. The first problem is a short-proof question. The second one involves derivation and analysis. The third problem is an application of least squares for approximating both linear and polynomial relationships between two variables.

Problem 1

Let

$$\mathbf{A} = \begin{bmatrix} \alpha & 1 & 0 \\ \beta & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

Find all values of α and β for which

1. $\mathbf{A}$ is singular.
2. $\mathbf{A}$ is symmetric.
3. $\mathbf{A}$ is positive definite.
4. $\mathbf{A}$ is strictly diagonally dominant.

Problem 2:

To deal with the efficiency of energy utilization of the larvae of the modest sphinx moth (Pachysphinx modesta), A researcher used the following data to determine a relation between W, the live weight of the larvae in grams, and R, the oxygen consumption of the larvae in milliliters/hour. For biological reasons, it is assumed that a relationship in the form of $R = bW^a$ exists between W and R.

Use the following given data points to approximate the unknowns a and b.

```r
W=c(0.017,0.025,0.020,0.020,0.025,0.087,0.111,0.085,0.119,0.233,0.174,0.211,0.171,0.210,0.783,
    1.110,0.999,1.290,1.320,1.350,1.740,3.020,3.040,3.340,1.690,4.090,4.280,4.290,5.480,2.750,
    5.450,4.580,5.300,4.830,5.960,4.680,5.530)

R=c(0.154,0.230,0.181,0.180,0.234,0.296,0.357,0.260,0.299,0.537,0.363,0.366,0.334,0.428,1.470,
    0.531,0.771,0.870,1.150,2.480,2.230,2.010,3.590,2.830,1.440,3.580,3.280,3.400,4.150,1.840,
    3.520,2.960,3.880,4.660,2.400,5.100,6.940)
```

[*Hint: R and W are nonlinear related. After taking logarithmic transformation on both sides of $R = aW^b$, $\ln W$ and $\ln R$ are linearly related and $\ln a$ and b are also linearly related.*]

Problem 3:

The cab data set contains data points that represent the duration (in minutes) from the midnight and number of customers picked up during the period. The data were collected in the 1970s in London. The objective of using this data set in the exam is to model the relationship between the duration and the customer count for traffic planning. The data set has been used in research and teaching. I made a copy of this data and stored it in the GitHub repository. You can use the following code to read the data to R.

```
cab = read.csv("https://raw.githubusercontent.com/pengdsci/MAT325/main/w11/cab-dataset.txt")
timeHours = cab$TimeMin/60     # x-coordinate: converted to hours
pickupCount = cab$PickupCount # y-coordinate
```

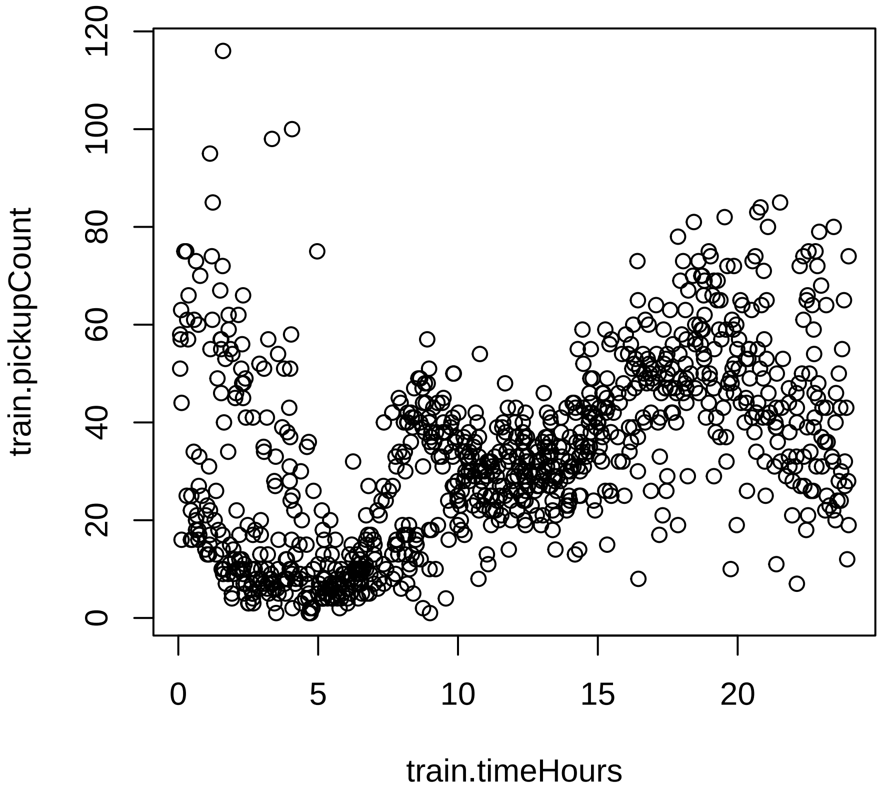
The data set will be randomly split into two subsets in the following:

```
valid.id = sample(1:length(timeHours), 0.25*length(timeHours), replace = FALSE)
valid.timeHours = timeHours[valid.id]
valid.pickupCount = pickupCount[valid.id]
###
train.timeHours = timeHours[-valid.id]
train.pickupCount = pickupCount[-valid.id]
```

Using **train.timeHours** and **train.pickupCount** to perform the following analysis. You are encouraged to use the code in the lecture notes or other R functions in your analysis.

1. Assume that **pickup count** and **duration** are linearly related. That is, $\text{train.pickupCount} = \alpha_0 + \alpha_0 \text{train.timeHours}$. Use the least square method to approximate this linear relationship.
2. Assume that **pickup count** and **duration** have a curve linear relationship: $\text{train.pickupCount} = \beta_0 + \beta_1 \text{train.timeHours} + \beta_2 \text{train.timeHours}^2 + \beta_3 \text{train.timeHours}^3$. Use the least square method to approximate this cubic polynomial.
3. Make a two-dimensional plot to visualize the relationship between the **duration (horizontal axis)** and **pickup-count (vertical axis)** and add the curves of the above approximated linear and cubic function to the same plot and comment on the goodness of the approximation.

```
plot(train.timeHours, train.pickupCount)
```



Bonus: Use the **valid.timeHours** and both linear and cubic polynomials to predict the *pickupCount*, denoted by **pred.pickupCount**, and calculate the mean (average) of squared error (**MSE**). Comment on the goodness of approximation using the MSE.