Homework Assignment Number Six Solutions

Problem (1) Single Variable Linear Regression

Perform a single-variable linear regression using the model

$$y = b_0 + b_1 x$$

(a) Report the mean value and standard deviation of the regression coefficients.
(b) Report the measure of fit.
Use the data in the file "file.hw06p01.txt" available on the website.
Note the first column in the data file is *y*. The second column is *x*.

Solution:

I used the code linreg1.m for linear regression with one independent variable.

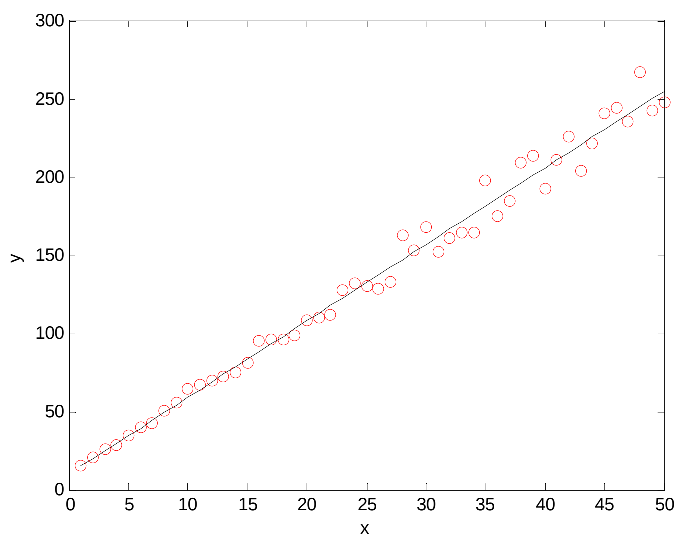
I wrote the small script hw06p01.m

```
clear all;

M = [15.10065279    1
20.4980324  2
25.56136963 3
...
242.1551242 49
247.5311911 50];

n = 50;
y = M(1:n,1);
x = M(1:n,2);

[b,bsd,MOF] = linreg1(x, y)
```

At the command line prompt, I executed the script

```
>> hw06p01
```

This generated the following output for the means, standard deviations and Measure of Fit.



b =
  10.074083107297966
   4.896473196043218

bsd =

```
   2.347470564943806
   0.080117982695322
```

MOF =
   0.987312123130968

The code also generated a plot.

Problem 2. Single Variable Polynomial Regression.
Perform a single-variable linear regression using the model

$$y = b_0 + b_1 x + b_2 x^2$$

(a) Report the mean value and standard deviation of the regression coefficients.
(b) Report the measure of fit.
Use the data in the file "file.hw06p02.txt" available on the website.
Note the first column in the data file is *y*. The second column is *x*.

Solution:

I used the code linregn.m for linear regression with one independent variable.

I wrote the small script hw06p02.m

```
clear all;

M = [28.60333775    2
61.67477585 4
113.330215  6
...
83236.71749 198
81985.43984 200];

n = 100;
y = M(1:n,1);
x(1:n,1) = M(1:n,2);
x(1:n,2) = x(1:n,1).*x(1:n,1);
[b,bsd,MOF] = linregn(2,x, y)
```

At the command line prompt, I executed the script

```
>> hw06p02
```

This generated the following output for the means, standard deviations and Measure of Fit.

```
b =
```

```
  86.579044451471418
   2.860222062350658
   2.015990010806434

bsd =   1.0e+02 *
   3.204973070484634
   0.073237990705217
   0.000351272052967

MOF =   0.998202555108216
```

The code does not generate a plot.

Problem 3. Multivariate Linear Regression
Perform a multivariate linear regression using the model

$$y = b_0 + b_1 x_1 + b_2 x_2$$

(a) Report the mean value and standard deviation of the regression coefficients.
(b) Report the measure of fit.
Use the data in the file "file.hw06p03.txt" available on the website.
Note the first column in the data file is *y*. The second column is x_1. The third column is x_2.

Solution:

I used the code linregn.m for linear regression with one independent variable.

I wrote the small script hw06p03.m

```
clear all;

M = [-0.981903982   2   10
-11.36111514    4   10
-17.92444528    6   10
...
188.825984  18  100
181.0279417 20  100];

n = 100;
y = M(1:n,1);
x(1:n,1) = M(1:n,2);
x(1:n,2) = M(1:n,3);
[b,bsd,MOF] = linregn(2,x, y)
```

At the command line prompt, I executed the script

```
>> hw06p03
```

This generated the following output for the means, standard deviations and Measure of Fit.

```
b =
 -19.426169012360049
  -4.968806248543629
   2.990195728881456

bsd =
   1.109461283774494
   0.066903032685564
   0.013380606537113

MOF =    0.998235942566897
```

The code does not generate a plot.

Problem 4. Reaction Rate Constants

Consider the isomerization reaction:

$$A \rightarrow B$$

The reaction rate is given by

$$rate = C_A k_o e^{-\frac{E_a}{RT}} \quad \text{[moles/liter/minute]}$$

where

concentration of A: C_A [moles/liter]
prefactor: k_o [1/minute]
activation energy for reaction: E_a [Joules/mole]
constant: $R = 8.314$ [Joules/mole/K]
temperature: T [K]

Determine the rate constants, k_o and E_a, from experimental data. The reaction is measured at a constant concentration of A, C_A =0.1 mol/liter, over a variety of temperatures. The rate is recorded. The rate as a function of temperature is given in tabular form in the file "file.hw06p04.txt" (containing 108 data points).

Convert the data into the form necessary for a linear regression.

$$\ln(rate) - \ln(C_A) = -\frac{E_a}{RT} + \ln(k_o)$$

This equation is of the form: $y = b_1 x + b_0$ where

$$y = \ln(rate) - \ln(C_A),\ b_1 = E_a,\ x = -\frac{1}{RT},\ \text{and}\ b_0 = \ln(k_o).$$

Solution:

I used the code linreg1.m for linear regression with one independent variable.

I wrote the small script hw06p04.m

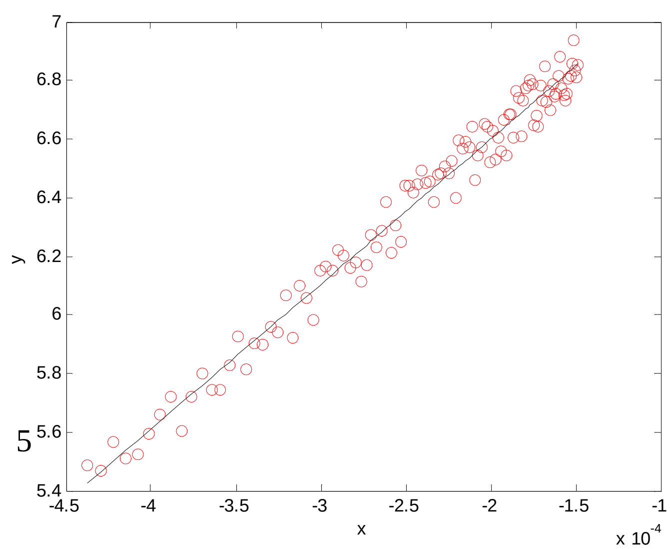
```
clear all;

%Temperature    rate of A loss
%K              mole/liter/min

M = [275    24.19186881
280 23.66394411
285 26.15820944
...
805 90.60688637
810 94.70386442];

R = 8.314; % J/mol/K
CA = 0.1; % mol/liter
n = 108;
for i = 1:1:n
    x(i) = -1.0/(R*M(i,1));
    y(i) = log(M(i,2)) - log(CA);
end
[b,bsd,MOF] = linreg1(x, y)
Ea = b(2)
ko = exp(b(1))
```

At the command line prompt, I executed the script

```
>> hw06p04
```

This generated the following output for the means, standard deviations and Measure of Fit.

```
b =   1.0e+03 *
   0.007590882856898
   4.952404881054963

bsd =
   0.018538702614459
  72.525898202990689

MOF =   0.977772149142637

Ea =
4.952404881054963e+03
```



```
ko =     1.980060852689931e+03
```

The activation energy was 4950 J/mol.
The rate constant was 1980 1/min.

The code also generated a plot.

Problem (5) Multivariate Nonlinear Optimization

Consider the rate equation

$$rate = k_o e^{-\frac{E_a}{RT+C}} \qquad \text{[mol/s]}$$

where

prefactor: k_o [mol/sec]
activation energy for reaction: E_a [Joules/mole]
non-Arrhenius parameter: C [Joules/mole]
constant: $R = 8.314$ [Joules/mole/K]
temperature: T [K]

Determine the rate constants, E_a, k_o and C, from experimental data. The rate as a function of temperature is given in tabular form in the file “file.hw06p05.txt”.

For initial guesses, use the knowledge that E_a should be on the order of 10,000 J/mol, k_o should be on the order of 100,000 mol/s and C should be on the order of 1000 J/mol.
Use whatever method you prefer. If you use the amoeba method, set the initial volume space to 50% of the initial guess and set your tolerance for both x and f to 10^{-8} or less. Use the RMS (root-mean-square error) as the objective function.

$$f_{obj} = \sqrt{\frac{1}{n_{data}} \sum_{i=1}^{n_{data}} \left(rate_i^{\exp} - rate_i^{\text{mod}} \right)^2}$$

Solution:

For this problem, I used the amoeba method. I modified the function in amoeba.m

```
function fobj = funkeval(x)
Ea = x(1);
```

```
k = x(2);
C = x(3);
R = 8.314;
datamat = [300  1.35E-02
320 2.97E-02
340 6.60E-02
360 1.14E-01
380 2.26E-01
400 3.44E-01
420 6.40E-01
440 1.10E+00
460 1.66E+00
480 2.21E+00
500 3.51E+00
520 4.97E+00
540 6.91E+00
560 9.26E+00
580 1.16E+01
600 1.59E+01];
ndata = max(size(datamat));
Tvec(1:ndata) = datamat(1:ndata,1);
rexp(1:ndata) = datamat(1:ndata,2);
for i = 1:1:ndata
   rmod(i) = k*exp(-Ea/(R*Tvec(i) + C));
end
fobj = 0.0;
for i = 1:1:ndata
    fobj = fobj + (rexp(i) - rmod(i))^2;
end
fobj = sqrt(fobj/ndata);
```

Per the instructions, I also changed line 51 of amoeba.m

```
lambda(1:np) = 0.5;
```

I executed the code with the following command

```
[f,x] = amoeba_hw06p05([10000,100000,1000],1.0e-8,1.0e-8);
```

This generated the following output

```
i =    1   1.0000000e+04   1.0000000e+05   1.0000000e+03 f =   1.2836808e+04
i =    2   1.5000000e+04   1.0000000e+05   1.0000000e+03 f =   4.9073912e+03
i =    3   1.0000000e+04   1.5000000e+05   1.0000000e+03 f =   1.9257600e+04
i =    4   1.0000000e+04   1.0000000e+05   1.5000000e+03 f =   1.5368448e+04
1    1.5000000e+04    1.0000000e+05   1.0000000e+03  4.9073912e+03  3.2659863e-01  1.1876858e+00
2    1.5000000e+04    0.0000000e+00   1.5000000e+03  5.9546926e+00  1.1775681e+00  1.9984508e+00
3    1.5000000e+04    0.0000000e+00   1.5000000e+03  5.9546926e+00  1.2000000e+00  1.9981454e+00
...
369  4.6619807e+04    7.4996156e+04   5.1492493e+02  1.4835916e-01  1.5121477e-08  9.9154342e-15
370  4.6619807e+04    7.4996156e+04   5.1492493e+02  1.4835916e-01  7.5607385e-09  9.9154342e-15
```

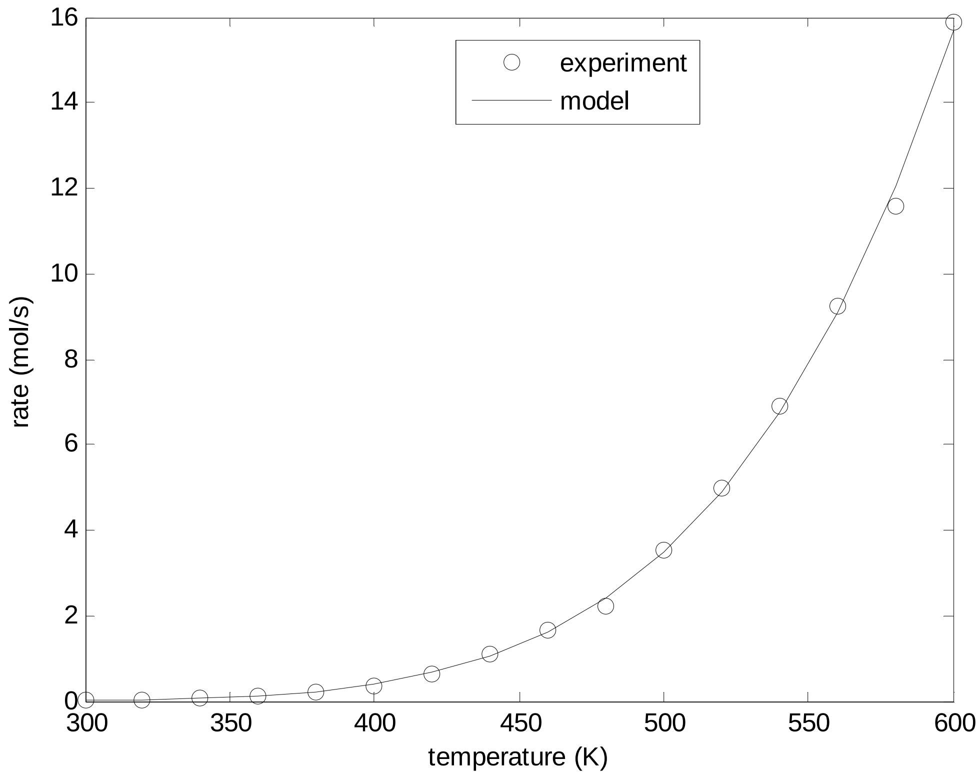
Therefore, we have the following optimized parameters

activation energy for reaction: $E_a = 46{,}600$ [Joules/mole]

prefactor: $k_o = 75{,}000$ [mol/sec]

non-Arrhenius parameter: $C = 515$ [Joules/mole]

A plot of the experimental rates and the model rates confirms the optimization was successful.



Problem (6) Fast Fourier Transforms

Using the Fast Fourier Transform, identify the frequencies present in the data file hw06p06.txt.

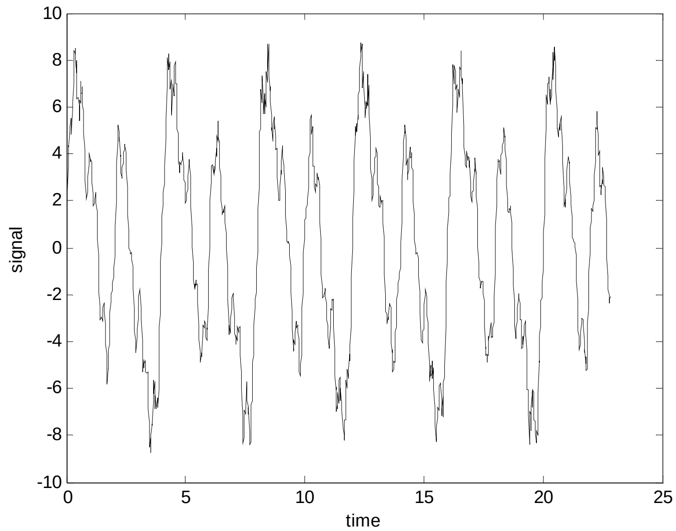
Solution

I wrote a short script and saved it as hw06p06.m.

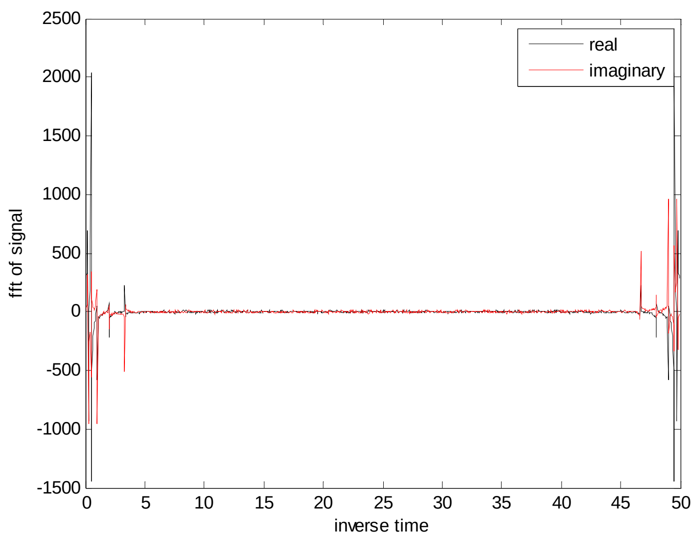
```
clear all;
close all;

%
%  fast fourier transform
%
datamat =[0 0
0.02    1.166832092
0.04    2.176137502
0.06    3.459186488
…
22.76   -2.036791092
22.78   -2.352011739
```

```
22.8    -2.213078457
22.82   -2.063030584];

ndata = max(size(datamat));

tvec(1:ndata) = datamat(1:ndata,1);
sig(1:ndata) = datamat(1:ndata,2);

figure(1)
plot(tvec,sig,'k-');
xlabel('time');
ylabel('signal');

dt = tvec(2) - tvec(1);
deni = 1.00/(ndata*dt);
kvec = zeros(ndata,1);
for i = 1:1:ndata
   ii = i - 1;
   kvec(i) = ii*deni;
end
fftsig = fft(sig);

figure(2)
plot(kvec,real(fftsig),'k-');
hold on;
plot(kvec,imag(fftsig),'r-');
hold off;
xlabel('inverse time');
ylabel('fft of signal');
legend('real','imaginary');

figure(3)
plot(kvec,real(fftsig),'k-');
hold on;
plot(kvec,imag(fftsig),'r-');
hold off;
xlabel('inverse time');
ylabel('fft of signal');
legend('real','imaginary');
axis([0 5 -1500 2000]);
```

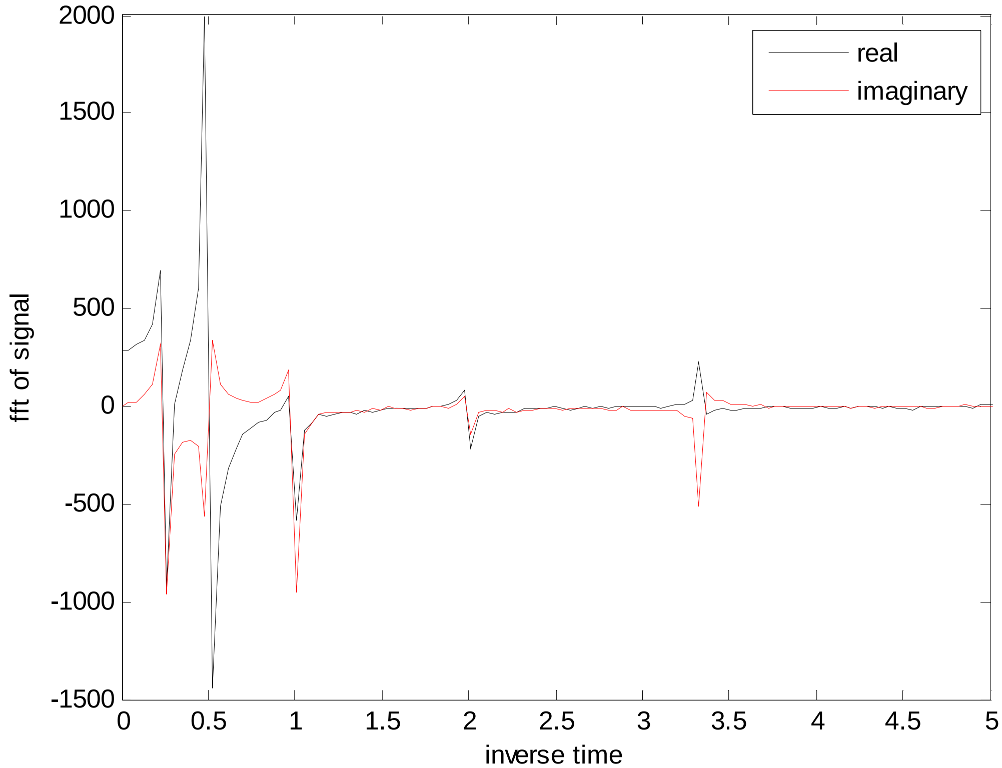
This script generated three plots.



This figure is a plot of the input data.



This figure is a plot of the output data from the fft command plotted verses k.



This figure is a close-up of the output. From this figure, we can observe that there are peaks at frequencies of 0.25, 0.5 1, 2, and 3.3. These correspond to periods of 0.3, 0.5, 1, 2 and 4.