

Two data sets are posted on Canvas. The first has an $n \times d = 1000 \times 50$ data matrix (X) "pred1.dat" with a 1000×1 response vector (y) in "resp1.dat." The second has a 1000×500 data matrix "pred2.dat" with a response vector in "resp2.dat." These data sets were generated according to the standard linear regression model.

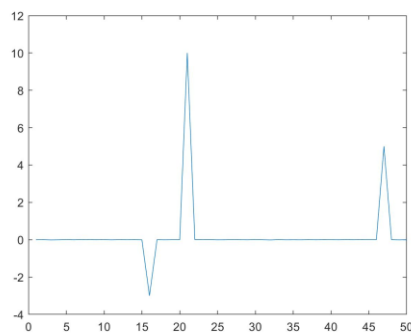
- For each data set, use the first half of the data to estimate values for W . Plot the estimated values for W , one from the first dataset and one from the other.
- For each data set, use your estimate of W on the 2nd half of the data set, to get your estimated response variables, \hat{Y} . Compute and report your total squared error.

a)

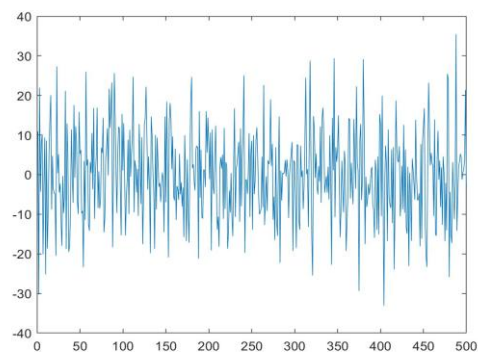
```
pred1 = importdata('pred1.dat');
resp1 = importdata('resp1.dat');
firstpred1 = pred1(1:500,:);
secpred1 = pred1(501:1000,:);
firstresp1 = resp1(1:500,1);
secresp1 = resp1(501:1000,1);
```

```
pred2 = importdata('pred2.dat');
resp2 = importdata('resp2.dat');
firstpred2 = pred2(1:500,:);
secpred2 = pred2(501:1000,:);
firstresp2 = resp2(1:500,1);
secresp2 = resp2(501:1000,1);
```

```
w1 = est(firstpred1, firstresp1);
w2 = est(firstpred2, firstresp2);
plot(w1)
```



```
plot(w2)
```



```
function b = est(pred, resp)
    b = inv(pred.' * pred) * pred.' * resp
end
```

b)

```
SSE1 = sum((secpred1 * w1 - secresp1).^2);
SSE2 = sum((secpred2 * w2 - secresp2).^2);
```

```
SSE1 = 5.7215
SSE2 = 3.2985e+07
```