Homework 2 - Question 3

```
data_1 = csvread('brca_reduced.csv');
Y_1 = data_1(:, end);
X_1 = data_1(:, 1:end-1);

%a)
disp("3a)")
fprintf("OLS fit \n")
X_1 = [ones(size(Y_1)),X_1];
B_ols_1 = regress(Y_1,X_1);
MSE_ols_1 = mean((Y_1 - X_1*B_ols_1).^2);
fprintf("MSE for clean data = %f \n",MSE_ols_1);


data_2 = csvread('brca_noisy.csv');
Y_2 = data_2(:, end);
X_2 = data_2(:, 1:end-1);
X_2 = [ones(size(Y_2)),X_2];
B_ols_2 = regress(Y_2,X_2);
MSE_ols_2 = mean((Y_2 - X_2*B_ols_2).^2);
fprintf("MSE for noisy data = %f \n", MSE_ols_2)
norm_ols = norm(B_ols_2-B_ols_1, inf);
fprintf("l-inf norm for noisy data = %f \n",norm_ols)

disp(" ")
%b)
disp("3b)")
fprintf("Robust fit using huber loss \n")
X_1 = data_1(:, 1:end-1);
B_robust_1 = robustfit(X_1,Y_1, 'huber');
MSE_robust_1 = mean((Y_1 - (X_1*B_robust_1(2:end) + B_robust_1(1))).^2);
fprintf("MSE for clean data= %f \n",MSE_robust_1)

X_2 = data_2(:, 1:end-1);
B_robust_2 = robustfit(X_2,Y_2, 'huber');
MSE_robust_2 = mean((Y_2 - (X_2*B_robust_2(2:end) + B_robust_2(1))).^2);
fprintf("MSE for noisy data = %f \n",MSE_robust_2)
norm_robost = norm(B_robust_2-B_robust_1, inf);
fprintf("l-inf norm = %f \n",norm_robost)
disp("On comparing a) and b) MSE of huber is more than OLS and huber results to more sparse coefficient matrix.");
disp(" ")
%c)
disp("3c)")
losses = ["cauchy", "talwar", "welsch"];
for i =1:length(losses)
    loss = losses(i);
    fprintf("Robust fit using %s loss \n", loss)
    B_robust_1 = robustfit(X_1,Y_1, loss);
    MSE_robust_1 = mean((Y_1 - (X_1*B_robust_1(2:end) + B_robust_1(1))).^2);
    fprintf("MSE for clean data = %f \n",MSE_robust_1);

    X_2 = data_2(:, 1:end-1);
    B_robust_2 = robustfit(X_2,Y_2, loss);
    MSE_robust_2 = mean((Y_2 - (X_2*B_robust_2(2:end) + B_robust_2(1))).^2);
    fprintf("MSE for noisy data = %f \n",MSE_robust_2)
    norm_robost = norm(B_robust_2-B_robust_1, inf);
    fprintf("l-inf norm = %f \n",norm_robost)
    disp(" ")
end


disp("We can see that in all the cases – OLS and Robust Regression, MSE calcuated for clean data is much lower than that calculated for noisy data
disp("MSE  for robust regression(in all 4 cases) is more than that of OLS for both noisy and clean data. The l-∞ norm of the difference between th
disp("regression coefficients, is much more in case of OLS. This shows that OLS is highly effected to outliers when compared to robust regression.
disp("Specifically for Cauchy, Talwar and Welsch loss, MSE is higher than Huber and same goes with sparsity of coefficients. This could indicate "
disp("they are more robust when compared to Huber");
```

```
3a)
OLS fit
MSE for clean data = 0.159327
MSE for noisy data = 10.422337
l-inf norm for noisy data = 2.239375

3b)
Robust fit using huber loss
MSE for clean data= 0.167658
MSE for noisy data = 13.604143
l-inf norm = 0.221294
On comparing a) and b) MSE of huber is more than OLS and huber results to more sparse coefficient matrix.

3c)
Robust fit using cauchy loss
MSE for clean data = 0.170836
MSE for noisy data = 14.490056
l-inf norm = 0.176621

Robust fit using talwar loss
MSE for clean data = 0.173959
MSE for noisy data = 14.742861
l-inf norm = 0.187621

Robust fit using welsch loss
MSE for clean data = 0.182239
```

```
MSE for noisy data = 14.698762
l-inf norm = 0.192031
```

We can see that in all the cases – OLS and Robust Regression, MSE calcuated for clean data is much lower than that calculated for noisy data MSE for robust regression(in all 4 cases) is more than that of OLS for both noisy and clean data. The l-∞ norm of the difference between the regression coefficients, is much more in case of OLS. This shows that OLS is highly effected to outliers when compared to robust regression. Specifically for Cauchy, Talwar and Welsch loss, MSE is higher than Huber and same goes with sparsity of coefficients. This could indicate they are more robust when compared to Huber

---

Published with MATLAB® R2018a