STA312 **Fall 2017**

Stanislav Volgushev **Homework 1, due Oct 6, 11:59 pm**

For problem 1 and 2, please submit an electronic solution via portal. Printed copies will not be accepted. Comment all of your programs in details. Problem 3 and 4 will be posted on portal, and the submission of solutions will be through portal directly. Note: residual plots will be explained in lectures on Monday, Sept 25

Problem 1 To make results reproducible, use the following code to generate data.

```
set.seed(42)
n = 100
x = runif(n,-1,1)
eps = rnorm(n,mean = 0, sd = .25)

xtest = - 1 + (1:1000)/500
epstest = rnorm(n,mean = 0, sd = .25)

y1 = x + eps
y2 = 2 - (x+1)^2 + eps
y3 = 1 + sin(4*x) + eps
```


1. `y1,...,y3` correspond to responses in a regression model. In each case, what is the true regression function f, what is the sample size, and what is the distribution of the error ε in that model?

2. For each of the response vectors `y1,...,y3`, run a simple linear regression with `x` as predictor (in each case, include an intercept). Provide the corresponding values for R^2 and the estimated regression coefficients. Make residual plots.

3. On each of the data sets, run K-nn regression with $K = 5, 10, 20$. For each of the 3 data sets provide plots showing the data points as well as 3 lines corresponding to the estimated K-nn regression functions with $K = 5, 10, 20$.

4. Use `xtest` and `epstest` to create 1000 test samples for each of the models given above. Use these test samples to compute the test error for the regression functions that you estimated in part 2 and 3. Comment on the results.

Problem 2 Solve Problem 12 on page 172 in the textbook Introduction to Statistical Learning.

Problem 3 and 4 will be posted on portal later this week. Those problems will allow you you to directly input your solutions to portal, no hand-in will be required.