**Assignment 4**

1. **That posterior median is optimal estimate under loss** (16 pts)
   1. Prove that the posterior median is optimal estimate under loss.
   2. Find the posterior median for the following posterior distribution
   3. Find the mean for the distribution defined in part **b** and discuss
   4. Discuss the difference between what you found in part **b** and **c**.

Note: For part b, you can start by finding the CDF of the distribution.

**2** **The decision rule for trading off FPs and FNs** (6 pts)

If , show that we should pick =1iff , where .

Note: and are the loss values for FN and FP errors.

**3** **Bootstrap Exercise** (20 pts)

Consider the following data listed below:

**X:** 3, 5, 7, 18, 43, 85, 91, 98, 100, 130, 230, 487

Carry out a nonparametric bootstrap analysis concerning the log of the sample mean as an estimator of the log of the population mean.

**(a)** Build the log the sample mean distribution for S=1000

**(b)** Estimate the mean of the log sample mean

**(c)** Estimate the standard error of the log sample mean

**(d)** Estimate the difference between the log sample mean and the log of the data mean

**(e)** Produce the 95% confidence interval for the log population mean

1. **Naïve Bayes, QDA, and LDA classifier (**18 pts**)**

Use 10-fold cross-validation and repeat the result for part c, d, e of **HW 3, Q6**

For the next two problems, we use the following dataset:

**Test Scores for General Psychology**  
  
The data (X1, X2, X3, X4) are for each student.  
X1 = score on exam #1  
X2 = score on exam #2  
X3 = score on exam #3  
X4 = score on final exam

1. **Linear regression (**20 pts**)**

Let’s assume X4 is the output variable - y.

1. Build a linear predictor based on X1, X2, or X3; and discuss which of these predictors give the lowest RSS. Show you work and prediction result
2. Build a linear predictor based on pairs of X1, X2, and X3; and discuss which pair gives the lowest RSS. Show you work and prediction result.
3. Build a linear predictor based on X1, X2, and X3. Provide the model parameters, show the residual error and RSS.
4. Discuss among all the models you built in part (a), (b), and (c) which one you pick as the proper model and why. You need to discuss it based on BIC and RSS.
5. **Logistic regression (**20 pts**)**

Let’s assume the threshold to pass the course is set at 160 based on the final exam (X4).

1. Build a linear predictor based on X1, X2, or X3; and discuss which of these predictors give the highest prediction accuracy. Show your work and discuss the prediction result.
2. Build a linear predictor based on pairs of X1, X2, and X3; and discuss which pair gives the highest prediction accuracy. Show your work and discuss prediction results.
3. Build a linear predictor based on X1, X2, and X3. Provide the model parameters, show the model’s prediction accuracy
4. Discuss among all the models you built in part (a), (b), and (c) which one you pick as the proper model and why. You need to discuss it based on BIC and performance.