STAT 598Z Final Exam

May, 2012

Time: 75 minutes

Your Name (Please print): ________________________

PUID (Please print): ________________________

Note:

1. This exam will contribute 20 points towards your final score
2. Use the provided scratch paper for your calculations
3. Take a print out of your code and attach it along with all relevant calculations
4. **Write your name clearly and legibly on all printouts**
5. Attempt as many problems as possible and explicitly state all your assumptions.
6. Show all intermediate steps for full credit. Python code must be clear and concise.
7. Any attempt at academic dishonesty (e.g. using a browser during the exam) will automatically result in 0 points.
8. Use of notes, books, laptops, cell phones, or any other aids (electronic or otherwise) is strictly prohibited. Turn off and put away your cell phone now!

Please sign below to indicate your agreement with the following honour code.

Honour code: I promise not to cheat on this exam. I will neither give nor receive any unauthorized assistance. I promise not to share information about this exam with anyone who may be taking it at a different time. I have not been told anything about the exam by someone who has already taken it.

Signature:________________________ **Date:**________________________

Questions	Possible Points	Actual Points
1	20/3	
2	20/3	
3	20/3	
Total	**20**	

Attempt any *three* out of the following five problems. All problems carry equal points.

Problem 1 Recall that the pdf of a Logistic random variable is

$$p(x|\lambda) = \frac{\lambda \exp(-\lambda x)}{(1+\exp(-\lambda x))^2} \text{ for } \lambda > 0.$$

- Derive the cdf of the above density. Show intermediate steps for full credit.
- Using pseudo-code describe a scheme for drawing samples from Logistic distribution.
- Implement your algorithm in Python.
- Generate 10,000 samples and plot their histogram.

Solution 1:

Problem 2 Suppose you observe the following data:

$$\{0, 1, 5, 7, 11\}$$

- Perform kernel density estimation using the following kernel

$$k(u) = \frac{1}{\sqrt{2\pi h^2}} \exp\left(-\frac{u^2}{2h^2}\right).$$

- Implement your algorithm in Python and plot the estimated density for $h \in \{0.1, 1.0, 10.0\}$.

Solution 2:

Problem 3 Suppose you are given the following two dimensional data:

$$\{x_1, x_2, x_3, x_4\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}.$$

Partition $\mathbb{R}^2$ into 4 regions such that all points in a region are closer to a given x_i than they are to any other x_j. In other words, if R_i is the region associated with x_i then

$$R_i = \{x \in \mathbb{R}^2 \text{ s.t. } \|x - x_i\| \leq \|x - x_j\| \quad \forall j \neq i\},$$

where $\|\cdot\|$ denotes the Euclidean norm. If such a partitioning is available to you then describe how you will use it to perform nearest neighbor classification. **Note:** You do not need to write any code for this problem.

Solution 3:

Problem 4 Draw 1000 samples from a mixture of normal distributions with the following parameters and plot the data

mean (μ)	Variance (σ)	proportion (π)
0	1	0.4
3	2	0.6

- Assume that σ values are given to you, while the sample proportion π and mean μ are latent variables. In the space below derive the Gaussian Mixture Model (GMM) updates for estimating π and μ. Note that this is different from the standard GMM derivation we did in the class. Here data is one dimensional. Furthermore, σ is observed but μ and π need to be estimated.
- Write a Python program which implements the GMM algorithm you derived above. Test your code on the random samples you generated above. Comment on your results.

Solution 4:

.

Problem 5

1. Prove that the following two functions are convex:
   - $f(x) = 4x^2 - 2x + 1$
   - $f(x) = \max(0, 1 - x)$
2. Plot the above functions for $x \in (-2, 2)$
3. Write a Python function which uses function values and gradients for minimizing a one dimensional convex function. Use your code to minimize $f(x) = 4x^2 - 2x + 1$. What is the minimum value and where is it attained?

Solution 5: