STAT 391
Homework 4
Out Saturday April 20, 2018
Due Tuesday May 1, 2018

mmp@stat.washington.edu

Problem 1 – Rescuing Rob

Rob is a robot that roams in the basement of the CSE building. The basement is fairly crowded and Rob has to detect and recognize several kinds of obstacles. For this he has a sonar that works in the following way: the sonar sends an ultrasound beam towards the obstacle, and then records the times at which it receives the returning echos. Then Rob's brain computes the distribution of the echos and matches it against it's "directory" of distribution signatures. There are three types of obstacles in Rob's catalog:

People When the "obstacle" is a person, Rob has to stop and let them pass by. A person has a sonar signature in the shape of a Gaussian (or normal) density.

Furniture These obstacles have to be detoured. Furniture appears on Rob's sonar as having a logistic CDF given by

$$F(x; a, b) = \frac{1}{1 + e^{-ax-b}} \tag{1}$$

$$f(x; a, b) = \frac{ae^{-ax-b}}{[1 + e^{-ax-b}]^2} \tag{2}$$

Trash When Rob encounters trash, he has to pick it up. Trash has a rather amorphous signature, and Rob has represented it with a kernel density estimator with Gaussian kernels.

This week, Rob walked too close to the Steam Powered Turing Machine[1] and a stream of boiling hot bits hit his memory circuits erasing the precious signature directory. Luckily, Rob still recalls that the last things he "saw" before the accident were: an undergraduate student, a boiler and an empty can of coke and he still has their signatures stored in the files `hw4-ugrad.dat, hw4-boiler.dat, hw4-coke.dat`. Help Rob recover his memory:

a. Estimate the parameters μ and σ^2 of the normal density that best fits the data in the file `hw4-ugrad.dat` by the Maximum Likelihood method. The data is in ASCII format, with spaces as separators.

b. Estimate the parameters a, b of the logistic density that best fits the data in the file `hw4-boiler.dat` by the Maximum Likelihood method. The data is in ASCII format, with spaces as separators. Make a plot of the log-likelihood of the data at each iteration.

Hint on doing the gradient ascent: Use the gradient formulas in the notes. Be careful chosing the step-size (and try a wide range of step-sizes to find a good one)! Plot the values of your parameters, and gradient, in addition to the log-likelihood. These plots are not required as part of the homework, but they are useful diagnostic tools. If you notice that the values oscillate, this is usually a sign that the step size is too large. If the values change too slowly, you can try increasing the step size. By the way, the step sizes that you try should be in a geometric not arithmetic progression (for example $\ldots\ 10^0\ 10^{-1}\ 10^{-2}\ 10^{-3}\ 10^{-4}\ 10^{-5}\ \ldots$ and not $\ldots 0.1\ 0.95\ 0.90\ 0.85\ \ldots$).

[1] `https://www.cs.washington.edu/building/art/SPTM`

c. Compute a kernel density estimate for the data in the file `hw4-coke.dat` using a Gaussian kernel with kernel width $h = 0.5$. *For this question, only write the formula of your density estimate, with h and n and the kernel replaced by the actual values.*

d. After having had his memory restored, Rob suddenly finds himself facing an unknown object whose signature is given in the file `hw4-unknown.dat`. Help Rob once more: tell him what is the object in front of him! For this, you need to compute the log-likelihood of the new data $\mathcal{D}_{unknown}$ under each of the three models. Print these log-likelihood values. The model that best predicts $\mathcal{D}_{unknown}$ (i.e gives it the highest log-likelihood) is Rob's guess of the identity of the object.

e. Make a plot of the densities evaluated in **a,b,c** and of the data $\mathcal{D}_{unknown}$ on the same graph. [Hint: to plot a density $f(x)$, choose a dense enough grid of points $x_{g1}, x_{g2}, \ldots x_{gN}$ and compute f at those points. Then plot $(x_{g1}, f(x_{g1}))$, $(x_{g2}, f(x_{g2})), \ldots$, etc.]

For this problem no proofs are required, but show the formulas that you use and the numerical results. In general, you should freely use the results from the lecture + course notes without proof. Submit the code that you used to solve this problem through the Assignments web page.

The task that you just performed for Rob is another example of *classification* – identifying an object as belonging to one of several *classes.* Rob's classes are people, furniture and trash. There are many methods of doing classification, not all of them statistical. The method that Rob is using, i.e comparing the likelihoods of the new data under the different class models, is a *likelihood ratio* method.

In the past robots used sonars for navigation, but these days they use lasers, GPS and cameras. They still use statistical methods to identify the shape of their environment.

Problem 2 – Maximum Likelihood with censored data

You are given samples $\{x_1, \ldots x_n\}$ from an exponential distribution with unknown parameter γ. But, by mistake, you store the data in the wrong format, which only preserves whether the data point was greater than 1 or not.

$$y_i \; = \; \begin{cases} 0 & \text{if } x_i \in [0, 1] \\ 1 & \text{if } x_i > 1 \end{cases} \qquad \text{for } i = 1 : n. \tag{3}$$

We say that the y_i observations are *censored* observations of the data x_i. With only the censored data $\{y_1, \ldots y_n\}$ you will estimate γ.

a. Write the probability that $y_i = 1$ as a function of γ.

b. Derive the expression of the log-likelihood $l(\gamma) = \ln P(y_{1:n}|\gamma)$ as a function of γ.

c. Maximize l w.r.t. γ and obtain the expression for γ^{ML}. [**Hint:** You may find it helpful to denote $\theta = e^{-\gamma}$ and use this variable in your maximization problem.]

d. Does this problem have sufficient statistics? How many and what are they?

[**e. Not graded**]*... but you are encouraged to answer or just think of this: Is the estimate of γ you are obtaining "better", "worse", "the same" as you would have obtained had you not lost the original data $x_{1:n}$? Can you give a formal meaning to "better"?*

[**f. Extra credit**] Sample $n = 100$ points from an exponential distribution with $\gamma = 1$, censor them, and estimate γ^{ML} from $x_{1:n}$ and γ^c from $y_{1:n}$. Compare the accuracy of the values obtained.

Note that if you only do this experiment once, the values γ^{ML}, γ^c are random, hence they can be (almost) anything. Therefore... For more credit, repeat the experiment N times, and display histograms of the N values γ^{ML}, γ^c obtained and comment on what you find. *This is a reasonable use of histograms.*

Problem 3 - Estimating h by cross-validation

For this problem, submit your code through the Assignments page link.

In this problem you will compute and plot a kernel density estimate of the corresponding densities f and

g given below

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

$$g(x) = \begin{cases} 4x & 0 \leq x \leq 0.5 \\ 4(1-x), & 0.5 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

a. Read in the training set D consisting of $n = 1000$ samples from F and validation set D_v of $m = 300$ samples from files `hw4-f-train.dat, hw4-f-valid.dat`. Use a kernel of your choice and then find the optimal kernel width h by cross-validation. For this, construct $f_h(x)$ the density estimated from D with kernel width h. Then compute the likelihood $L_v(h)$ of the data in D_v under f_h. Also compute $L(h)$, the likelihood of the training set D under f_h. Repeat this for several values of h and plot $L_v(h)$ and $L(h)$ as a function of h on the same graph. (Suggested range of h: 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5).

Let h^* be the h that maximizes $L_v(h)$. Make a plot of $f_{h^*}(x)$ (by, for instance, computing the $f_{h^*}(x)$ values on a grid $x = -0.5, -0.49, -0.48, \ldots 1.49, 1.5$). Plot the true $f(x)$ on the same graph.

The homework you hand in should contain: the equation of the chosen kernel, the formula(s) you used for f_h for the chosen kernel, the formula(s) you used to compute $L_v(h)$ and $L(h)$ and the required graphs. It is OK to replace likelihoods with log-likelihoods in the plots and equations.

b. Do the same for G and g, reading data from the files `hw4-g-train.dat, hw4-g-valid.dat`.

c. Compare the optimal h's and the quality of the plots in **a, b**. Which of the densities looks easier to approximate? Which of the optimal kernels widths is larger, the one used for f or the one used for g? Can you suggest an explanation why?