

# Homework 1(A) for STA410/2102

Due back before 11pm on October 20, 2016

**Note:** Graduate students shall solve all problems. Undergrads shall solve problems 1, 2 and 3. You will soon receive instructions on how to submit the solutions for grading.

## Problem 1

The file `copula-hw1.txt` contains 200 samples for the bivariate random variable  $(X, Y)$  whose density is

$$f(X, Y | \mu_1, \mu_2, \theta) = \frac{\phi(X - \mu_1)\phi(Y - \mu_2)\theta(1 - e^{-\theta})e^{-\theta[\Phi(X - \mu_1) + \Phi(Y - \mu_2)]}}{[1 - e^{-\theta} - (1 - e^{-\theta\Phi(X - \mu_1)})(1 - e^{-\theta\Phi(Y - \mu_2)})]^2},$$

where  $\phi(x)$  is the density of a standard normal distribution  $\phi(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$  and  $\Phi(x) = \int_{-\infty}^x \phi(t)dt$  is the cumulative distribution function of a standard normal distribution.

Using the provided data, compute the maximum likelihood estimator for the model parameters  $\mu_1 \in \mathbf{R}$ ,  $\mu_2 \in \mathbf{R}$  and  $\theta \in \mathbf{R}$ .

**Problem 2** The file `antithetic-boot.txt` contains an iid sample of size  $n = 64$  from the distribution of  $Y$ . Using the combinatorial optimization method based on local search, find permutations  $\tau_1$  and  $\tau_2$  such that

$$\sum_{j=1}^{64} [Y_j Y_{\tau_1(j)} + Y_j Y_{\tau_2(j)} + Y_{\tau_1(j)} Y_{\tau_2(j)}]$$

is minimum.

**Problem 3** Thirteen chemical measurements were carried out on each of 178 wines from three regions of Italy. These data are available from the file `wine.txt`. Using one of the optimization methods discussed in this course, partition the wines into three groups for which the total of the within-group

sum of squares is minimal. Specifically, suppose that the super-index  $(h)$  marks the cluster assigned to the  $i$ th vector of 13 measurements,  $\vec{x}_i$ . The within-group sum of squares for cluster  $h$  is then  $\sum_{i=1}^{n_h} \|\vec{x}^{(h)} - \bar{x}^{(h)}\|^2$  where  $\bar{x}^{(h)}$  is the average of all measurements in cluster  $h$ , and for  $a \in \mathbf{R}^{13}$  we define  $\|a\|^2$  is  $\sum_{i=1}^{13} a_i^2$ . Note that this is a search problem of size  $3^p$  where  $p = 178$ .

**Problem 4**<sup>1</sup> Using genetic algorithms, solve Problem 2.

**Problem 5**<sup>2</sup> The baseball.dat presents data on baseball player salaries along with a number of covariates. Find which covariates are included in the model when the objective function is the penalized least squares:

$$PLS(\beta_{MLE}) = \sum_{i=1}^n (y_i - x_i^T \beta_{MLE})^2 + \lambda \|\beta_{MLE}\|_4^*$$

where for any  $\beta = (\beta_0, \dots, \beta_{p-1}) \in R^p$

$$\|\beta\|_4^* = \left( \sum_{j=1}^{p-1} \beta_j^4 \right)^{1/4}.$$

(note that the intercept  $\beta_0$  is not included in the penalty term. )

Study the difference between the final models when  $\lambda = 1, 5, 10$ .

---

<sup>1</sup>ONLY for Graduate Students

<sup>2</sup>ONLY for Graduate Students