EE 641 Midterm Exam
October 24, Fall 2016

Name: Solution

Instructions

The following is an in-class closed-book exam.

- This exam contains 3 problems worth a total of 100 points.
- You may not use any notes, textbooks, or calculators.
- You are allowed up to 55 minutes to complete the exam.

Good luck.

Problem 1. (30pt)
Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables with distribution

$$P\{X_i = k\} \;=\; \pi_k$$

where $\sum_{k=1}^m \pi_k = 1$. Compute the ML estimate of the parameter vector $\theta = [\pi_1, \cdots, \pi_m]$. (Hint: You may use the method of Lagrange multipliers to calculate the solution to the constrained optimization.)

Solution

$$P(X_i = x_i) = \pi_{x_i} = \prod_{k=1}^{m} \pi_k^{\delta(x_i - k)}$$

$$P(x_1, x_2, \cdots, x_n | \theta) = \prod_{i=1}^{n} \prod_{k=1}^{m} \pi_k^{\delta(x_i-k)} = \prod_{k=1}^{m} \prod_{i=1}^{n} \pi_k^{\delta(x_i-k)}$$

$$= \prod_{k=1}^{m} \pi_k^{N_k}$$

where $N_k = \sum_{i=1}^{n} \delta(x_i - k)$ i.e. N_k is the number of times output k is observed.

Since the parameters $\pi_1, \pi_2, \cdots, \pi_m$ must add to 1, so we use Lagrange multipliers.

$$L(\theta, \lambda) = \log p(x_1, x_2, \cdots, x_n | \theta) + \lambda \left(\sum_{k=1}^{m} \pi_k - 1 \right)$$

$$= \sum_{k=1}^{m} N_k \log \pi_k + \lambda \left(\sum_{k=1}^{m} \pi_k - 1 \right)$$

Maximize $L(\theta, \lambda)$ w.r.t $[\theta, \lambda]$

$$\left. \frac{\partial L(\theta, \lambda)}{\partial \pi_k} \right|_{\substack{\pi_k = \hat{\pi}_k \\ \lambda = \hat{\lambda}}} = \frac{N_k}{\hat{\pi}_k} + \hat{\lambda} = 0 \qquad \text{—————} \qquad (1)$$

$$\left.\frac{\partial}{\partial \lambda} L(\theta, \lambda)\right|_{\substack{\bar{\pi}_k = \hat{\pi}_k \\ \lambda = \hat{\lambda}}} = \sum_{k=1}^{m} \hat{\pi}_k - 1 = 0$$

Solving eq ① and ② simultaneously, we get

$$\frac{N_k}{\hat{\pi}_k} + \hat{\lambda} = 0$$

$$\Rightarrow N_k = -\hat{\lambda}\,\hat{\pi}_k$$

$$\Rightarrow \sum_{k=1}^{m} N_k = -\hat{\lambda} \sum_{k=1}^{m} \hat{\pi}_k \quad \because \text{using eq ②}$$

$$\Rightarrow \hat{\lambda} = -n \quad \because \sum_{k=1}^{m} N_k = n$$

$$\Rightarrow \hat{\pi}_k = \frac{N_k}{n} \quad \because \text{putting value of } \hat{\lambda} \text{ in eq ①}$$

$\hat{\pi}_k$ is interpreted as the fraction of times k^{th} output is observed

Note:

It can be shown using the Hessian of $L(\theta, \lambda)$ that the above solution is the global maximizer of the likelihood function under the constraint assuming all N_k are positive.

Problem 2. (35pt)

Let X_s be a zero-mean GMRF on a finite general lattice $s \in S$. Let X be a vector of dimension $N = |S|$ containing all the elements of X_s in some fixed order, and denote the inverse covariance of X as

$$B = \left(\mathbb{E}[XX^t]\right)^{-1} .$$

a) Write an expression for $p(x)$, the PDF of X in terms of B.

b) If ∂s denotes the neighborhood system of the MRF, then show that if $r \notin \partial s$ and $r \neq s$, then $B_{r,s} = B_{s,r} = 0$.

c) Show that we can define a valid (but possibly different) neighborhood system for this GMRF as

$$\partial s = \{r \in S : B_{r,s} \neq 0 \text{ and } r \neq s\} .$$

Solution

(a) $$p(x) = \frac{|B|^{1/2}}{(2\pi)^{N/2}} \exp\left\{ -\frac{1}{2} x^t B x \right\}$$

(b) For an MRF, $p(x_s | x_{r \neq s}) = p(x_s | x_{\partial s})$

Now

$$p(x_s | x_{r \neq s}) \propto \exp\left\{ -\frac{1}{2} x^t B x \right\}$$

$$\Rightarrow p(x_s | x_{r \neq s}) \propto \exp\left\{ -\frac{1}{2} \left(x_s^2 B_{s,s} + 2x_s \sum_{\substack{r \in S \\ r \neq s}} B_{s,r} x_r \right) \right\}$$

Completing the square, we get

$$p(x_s | x_{r \neq s}) = \frac{1}{z} \exp\left\{ \frac{-1}{2\sigma_s^2} \left(x_s + \sigma_s^2 \sum_{\substack{r \in S \\ r \neq s}} B_{s,r} x_r \right)^2 \right\}$$

where $\sigma_s^2 = B_{s,s}^{-1}$

Since we know that $p(x_s | x_{r \neq s}) = p(x_s | x_{\partial s})$

Therefore $B_{s,r} = B_{r,s} = 0$ for $r \notin s$ and $r \neq s$.

(c) $p(x_s \mid x_{r \neq s}) = \frac{1}{z} \exp\left\{ \frac{-1}{2\sigma_s^2} \left(x_s + \sigma_s^2 \sum_{\substack{r \in S \\ r \neq s}} B_{s,r}\, x_r \right)^2 \right\}$

Now if we let $\partial s = \{ r \in S : B_{r,s} \neq 0 \text{ and } r \neq s \}$, then

$$p(x_s \mid x_{r \neq s}) = \frac{1}{z} \exp\left\{ \frac{-1}{2\sigma_s^2} \left(x_s + \sigma_s^2 \sum_{r \in \partial s} B_{s,r}\, x_r \right)^2 \right\}$$

$$\Rightarrow p(x_s \mid x_{r \neq s}) = p(x_s \mid x_{\partial s})$$

The above equation is the definition for an MRF. Therefore, the given set $\partial s = \{ r \in S : B_{r,s} \neq 0 \text{ and } r \neq s \}$ gives an alternative neighborhood system.

<u>Note:</u> This neighborhood system can be different from the earlier neighborhood specification ∂s^{old} if $\exists\, r$ s.t. $r \in \partial s^{old}$ and $B_{s,r} = 0$.

Problem 3. (35pt)
Consider the function

$$f(x) = |x - x_r|^{1.1} ,$$

for $x \in \mathbb{R}$.

a) Sketch a plot of $f(x)$ when $x_r = 1$.

b) Sketch a good surrogate function, $f(x; x')$, for $x_r = 1$ and $x' = 2$.

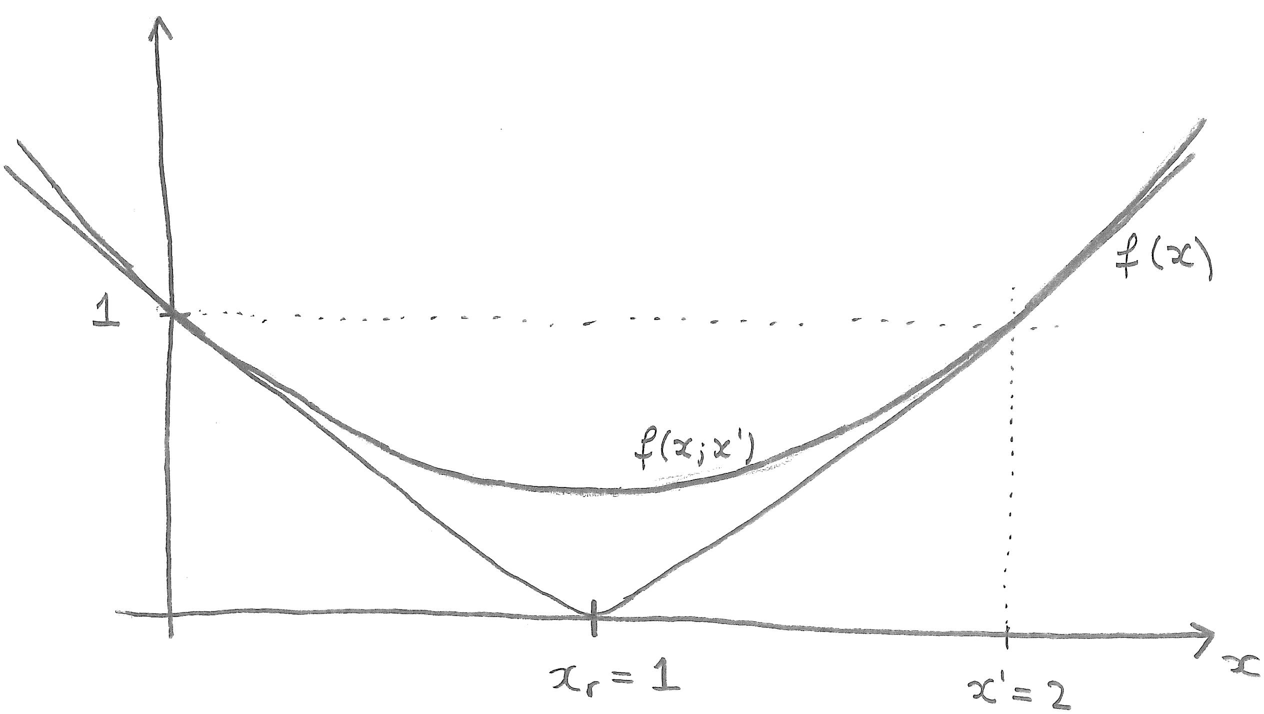
c) Determine a general expression for the surrogate function $f(x; x')$ that works for any value of x_r and x'.

d) Assuming the objective is to minimize the expression

$$f(x) = \sum_{r \in \partial s} |x - x_r|^{1.1} ,$$

for $x \in \mathbb{R}$, specify an iterative algorithm in terms of the surrogate function $f(x; x')$ that will converge to the global minimum of the function.

(a) and (b)



(c) Construction of surrogate function (for $x' \neq x_r$).

$$f(x; x') = \frac{a}{2}(x - x_r)^2 \quad \because \text{ using symmetric bound.}$$

Pick a to match the first derivatives of $f(x)$ and $f(x; x')$.

$$\frac{d}{dx} f(x) = 1.1\, |x - x_r|^{0.1} \operatorname{sgn}(x - x_r) \quad \text{———} \quad (1).$$

$$\frac{d}{dx} f(x; x') = a(x - x_r) \quad \text{———} \quad (2)$$

Equate eq (1) and eq (2) at $x = x'$, $(x' \neq x_r)$

$$\Rightarrow \quad a = \frac{1.1\, |x' - x_r|^{0.1} \operatorname{sgn}(x' - x_r)}{(x' - x_r)}$$

$$\Rightarrow \quad a = 1.1\, |x' - x_r|^{-0.9} \quad \because \frac{\operatorname{sgn}(\Delta)}{\Delta} = \frac{1}{|\Delta|}$$

when $x' = x_r$, $a = f''(x_r) = \infty$

So the surrogate for $x' \neq x_r$ is

$$f(x; x') = \frac{1.1\, |x' - x_r|^{-0.9}}{2} \cdot (x - x_r)^2.$$

(d) Minimization of $f(x) = \sum_{r \in \partial s} |x - x_r|^{1.1}$

Using additivity of surrogate functions, we form a surrogate function

$$f(x; x') = \sum_{r \in \partial s} \frac{1.1 \, |x' - x_r|^{-0.9}}{2} (x - x_r)^2 \qquad (3)$$

To minimize the surrogate in eq (3), we set

$$\left. \frac{d}{dx} f(x; x') \right|_{x = x^*} = 0$$

$$\Rightarrow \quad x^* = \frac{\sum_{r \in \partial s} |x' - x_r|^{-0.9} \, x_r}{\sum_{r \in \partial s} |x' - x_r|^{-0.9}}$$

Thus, we have a simple iterative algorithm

Repeat Untill Convergence

$$\left\{ \quad x^{(k+1)} \leftarrow \frac{\sum_{r \in \partial s} |x^{(k)} - x_r|^{-0.9} \, x_r}{\sum_{r \in \partial s} |x^{(k)} - x_r|^{-0.9}} \quad \right\}$$