Lecture 6: Sensor Selection

Yilin Mo

July 2, 2015

1 Problem Formulation

Let $x \in \mathbb{R}^n$ be the state, such that $x\,\mathcal{N}(0, \Sigma)$. $y \in \mathbb{R}^m$ is the sensor measurement, where y_i is the measurement from the ith sensor, such that

$$y_i = a_i x + v_i.$$

Assume that $x, v_1, \dots, v_m$ are all linearly independent and $v_i \sim \mathcal{N}(0, 1)$. (without loss of generality we can assume $\mathbb{E}v_i^2 = 1$, why?)

Let γ_i be a binary variable, such that $\gamma_i = 0$ if the sensor is not selected. $\gamma_i = 1$ if the sensor is selected.

Hence, the estimation covariance is given by

$$P = \left(\Sigma^{-1} + \gamma_i \sum_i a_i a_i^T\right)^{-1}.$$

Possible objective functions:

- $\text{tr}(P)$
- $\log\det(P) = -\log\det(\Sigma^{-1} + \gamma_i \sum_i a_i a_i^T)$, $\log\det(X)$ is a concave function with respect to $X \in \mathbb{S}_+^n$.
- Why not P?

Possible constrains:

- sensor i is selected only when sensor j is selected: $\gamma_i \leq \gamma_j$
- sensor i and j can not both be selection: $\gamma_i + \gamma_j \leq 1$
- number constraints: $\sum_i \gamma_i \leq l$.
- budget constraints: $\sum_i w_i \gamma_i \leq l$.

The main difficulty is that γ_i is discrete. In general, sensor selection problem is NP-hard.

2 Convex Relaxation Based Approach

We consider the following problem:

$$\begin{aligned} \underset{\gamma_1,\ldots,\gamma_m}{\text{minimize}} \quad & \operatorname{tr}(P^{-1}) \\ \text{subject to} \quad & \sum_i \gamma_i \le l \\ & \gamma_i \in \{0,1\} \end{aligned}$$

A standard way to deal with binary variables is to relax it to a real number.

Relaxed Problem:

$$\begin{aligned} \underset{\gamma_1,\ldots,\gamma_m}{\text{minimize}} \quad & \operatorname{tr}(P^{-1}) \\ \text{subject to} \quad & \sum_i \gamma_i \le l \\ & 0 \le \gamma_i \le 1 \end{aligned}$$

Let the solution of the relaxed problem be $\gamma_1^*, \ldots, \gamma_m^*$. We can quantize the real $\gamma_1^*, \ldots$ to a binary $\gamma_1, \ldots, \gamma_m$. Let the solution of the original problem be $\gamma_1^o, \ldots, \gamma_m^o$. Hence

$$\operatorname{tr}(P^{-1}(\gamma^*)) \le \operatorname{tr}(P^{-1}(\gamma^o)) \le \operatorname{tr}(P^{-1}(\gamma)).$$

As a result, we have an estimate on how good our approximation is.

3 Submodularity

Let S be a set and $f : 2^S \to \mathbb{R}$ be a function.

The function is called monotone if for all $A \subseteq B \subseteq S$,

$$f(A) \le f(B).$$

The function is called submodular if for any $A, B \subseteq S$,

$$f(A \cap B) + f(A \cup B) \le f(A) + f(B).$$

Equivalently, if $A \subseteq B$ and $i \notin B$, then

$$f(A \cup \{i\}) - f(A) \ge f(B \cup \{i\}) - f(B).$$

Examples of monotone and submodular functions:

- $f(S) = \sqrt{|S|}$.
- Linear function: assume that every $i \in S$ has a weight w_i, $f(A) = \sum_{i \in A} w_i$.

- Coverage function: Let X be a weighted set and $X_i \subseteq X$ for $i = 1, \ldots, n$. Let $S = \{1, \ldots, n\}$. Then the following function is submodular:

$$f(A) = \sum_{x \in \bigcup_{i \in A} X_i} w(x).$$

Let $Y = [y_1, \ldots, y_m]$. Let $A = \{i_1, \ldots, i_l\} \subseteq \{1, \ldots, m\}$ and $Y_A = [y_{i_1}, \ldots, y_{i_l}]$. Define the mutual information

$$I(X; Y_A) = h(X) - h(X|Y_A).$$

In general, this function is monotone but not increasing. Counterexample: y_1, y_2 are Bernoulli with $P(y_i = 1) = 0.5$ and $X = y_1 \text{XOR} y_2$. Then X is independent of y_1 and X is independent of y_2. Hence,

$$f(\emptyset) = f(\{1\}) = f(\{2\}) = 0,\ f(\{1, 2\}) = 1.$$

However, if y_is are conditionally independent, i.e.,

$$P(y_1, \ldots, y_m|X) = \prod_i P(y_i|X),$$

then the mutual information is monotone submodular.

3.1 Maximizing a Monotone Submodular Function

Consider the following problem:

$$\begin{array}{ll} \underset{A \subset S}{\text{maximize}} & f(A) \\ \text{subject to} & |A| \leq l \end{array}$$

Define the optimal A for the above problem to be A^*. This problem is in general NP-hard. However, we have a greedy algorithm:

1. Let $A_0 = \emptyset$.
2. Find $s = \arg\max_{s \in S} f(A_i \cup \{s\})$. Get $A_{i+1} = A_i \cup \{s\}$.
3. Iterate until we get A_k.

Theorem 1. *Let f be a non-negative monotone submodular function, then*

$$f(A_l) \geq (1 - e) f(A^*)$$

Proof. Let $A^* = \{s_1, \ldots, s_k\}$. Hence,

$$f(A^*) \leq f(A^* \cup A_i) \leq f(A_i) + \sum_{s_i \in A^*} (f(A_i \cup \{s_i\}) - f(A_i)) \leq f(A_i) + k(f(A_{i+1}) - f(A_i)),$$

which is equivalent to

$$f(A^*) - f(A_{i+1}) \leq \left(1 - \frac{1}{k}\right)(f(A^*) - f(A_i)).$$

Hence,

$$f(A^*) - f(A_k) \leq \left(1 - \frac{1}{k}\right)^k (f(A^*) - f(A_0)) \leq \frac{1}{e} f(A^*).$$

The last inequality holds since $1 - 1/k \leq e^{-1/k}$. □

3.2 Greedy Sensor Selection

For a Gaussian distribution $\mathcal{N}(0, \Sigma)$, the entropy is given by

$$\frac{k}{2}(1 + \log(2\pi)) + \frac{1}{2}\log\det\Sigma,$$

where k is the dimension of the Gaussian variable. For our case, the mutual information between x and Y_A is

$$\frac{1}{2}\log\det\Sigma - \frac{1}{2}\log\det P.$$

Hence, we can solve the following problem using the greedy algorithm:

$$\begin{array}{ll} \underset{\gamma_1,\ldots,\gamma_m}{\text{minimize}} & \log\det P \\ \text{subject to} & \sum_i \gamma_i \leq l \\ & \gamma_i \in [0, 1] \end{array}$$