Lecture Note of STA250

Chun-Jui Chen

November 4,2013

1 Python

- 1. For python 2, the most used version is 2.6 and 2.7
- 2. For python 3, the most used version is 3.2 and 3.3
- 3. Python is good for text processing
- 4. Two ways to run python.
  i, under the environment of python, use the interactive mode.
  ii, Executive python scrip.
- 5. For dealing with string:
  baz="Class"
  bar="Hello"
  foo2=bar+" "+baz
- 6. Indentation is important in python:
- 7. python uses zero indexing.
- 8. Format can only be converted by following order: Str $\rightarrow$ Float $\rightarrow$ int

2 Bayes

Bayes recap: Assume $X_1, ..., X_n$ follow i.i.d. $N(\mu, \sigma^2)$, σ^2 known and $\mu \sim N(\mu_0, \sigma_0^2)$, then

$$\mu|\vec{X} \sim N(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}})$$

Assume that we have general likelihood $p(\vec{X}|\theta)$ and prior $p(\theta)$, thus the posterior will be proportional to:

$$p(\theta|\vec{X}) \propto p(\theta)p(\vec{X}|\theta)$$

If $p(\theta|\vec{X})$ has no closed form, use MCMC.

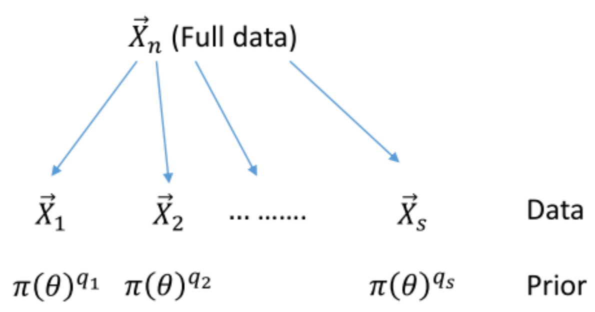
3 Big Data

For big data, applying MCMC directly to the full dataset is not computationally feasible.

Idea: "chunk" full dataset into a series of smaller datasets, sample from posterior of each smaller dataset and combine results.
Assume that $X_i's$ are conditionally independent given θ

$$\prod_{j=1}^{s} p(\theta)^{1/s} p(\vec{X}_j|\theta)$$

where $\vec{X}_j$ is the j-th subset of data and s is the total number of subsets.



where $\sum q_j = 1$, $q_j > 0\ \forall j$ and $\cup_j \vec{X}_j = \vec{X}$, $\cap_j \vec{X}_j = \emptyset$. Aside,

$$\mu \sim N(\mu_0, \sigma_0^2)$$
$$p(\mu)^{q_j} \propto \exp\{-\frac{q_j}{2\sigma_0^2}(\mu - \mu_0^2)\}$$

4 Consensus Monte Carlo

Assume that sample $\theta_j^{(t)}$ from model m_j (i.e. $p_{m_j}(\theta|\vec{X}_j) \propto \pi(\theta)^{q_j} p(\vec{X}_j|\theta)$). Combine to

$$\theta^{(t)} = \frac{\sum_{j=1}^{s} w_j \theta_j^{(t)}}{\sum_{j=1}^{s} w_j}$$

where $w_j > 0$ are known weight. For weights, it turns out that a good good choice is: $w_j = var^{-1}(\theta|\vec{X}_j)$.

Example. If $X_1, ..., X_n|\mu$ follow i.i.d. $N(\mu, \sigma^2)$ and $\mu \sim N(\mu_0, \sigma_0^2)$ and chunk data into $\vec{X}_j = \{X_{j_1}, X_{j_2}, ..., X_{j_k}\}$. Where $\{j_1, ..., j_k\}$ are indices corresponding to the subset j.
Model j:

$$X_{j_1}, X_{j_2}, ..., X_{j_k}|\mu \sim_{iid} N(\mu, \sigma^2)$$

$$\mu \sim N(\mu_0, \frac{\sigma_0^2}{q_j})$$

Posterior for Model j:

$$\mu_j|\vec{X}_j \sim N\left(\frac{\frac{q_j\mu_0}{\sigma_0^2} + \frac{\sum x_j}{\sigma^2}}{\frac{q_j}{\sigma_0^2} + \frac{k_j}{\sigma^2}}, \frac{1}{\frac{q_j}{\sigma_0^2} + \frac{k_j}{\sigma^2}}\right)$$

Combine $\mu = \frac{\sum w_j \mu_j}{\sum w_j}$, $w_j = \frac{\varepsilon_j}{\sigma_0^2} + \frac{k_j}{\sigma^2}$ and $\sum w_j = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$. Then, we get

$$\begin{aligned} Var(\mu) &= \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^2} \sum_j w_j^2 Var(\mu_j) \\ &= \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \end{aligned}$$

$$\begin{aligned} E(\mu) &= \frac{\sum_j w_j E[\mu_j]}{\sum w_j} = \frac{\sum_j \frac{q_j\mu_0}{\sigma_0^2} + \frac{\sum\sum x_{k_j}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \\ &= \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_j}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \end{aligned}$$

Here, evenly chunk data is suggested.
Suppose n is large and the sub datasets are also large then we can get the posterior approximates

normal distribution. (MC will still work well)