Robust Inversion, Dimensionality Reduction, and Randomized Sampling

——Course Presentation

Shan You, Kai Zheng, Cong Fang

December 15, 2014

Outline

1. Why dimension reduction
2. Random Sampling
   - Data Averaging
   - Sampling with and without replacement

$$\text{Recall}: \min \phi(x) = \frac{1}{m}\sum_{i=1}^{m} \phi_i(x),$$
$$\text{where } \phi_i(x) = \rho(r_i(x)) = \rho(d_i - F_i(x)q_i)$$

Disadvantage: Expensive Computation

How to solve? Stochastic Gradient Descent

$$\text{Recall}: \min \phi(x) = \frac{1}{m}\sum_{i=1}^{m}\phi_i(x),$$
$$\text{where } \phi_i(x) = \rho(r_i(x)) = \rho(d_i - F_i(x)q_i)$$

Disadvantage: Expensive Computation

How to solve? Stochastic Gradient Descent

$$\text{Recall}: \min \phi(x) = \frac{1}{m}\sum_{i=1}^{m}\phi_i(x),$$
$$\text{where } \phi_i(x) = \rho(r_i(x)) = \rho(d_i - F_i(x)q_i)$$

Disadvantage: Expensive Computation

How to solve? Stochastic Gradient Descent

Outline

1. Why dimension reduction
2. Random Sampling
   - Data Averaging
   - Sampling with and without replacement

When loss function is square function,

$$\phi(x) = \frac{1}{m}\mathrm{tr}(R(x)^T R(x))$$

where $R(x) = [r_1(x), r_2(x), \ldots, r_m(x)]$

'New Data': $\tilde{d}_j = \sum_{i=1}^{m} w_{ij} d_i, \tilde{q}_j = \sum_{i=1}^{m} w_{ij} q_i, j = 1, \ldots, s$
$\tilde{r}_j(x) = \tilde{d}_j - F(x)\tilde{q}_j, R_W(x) := [\tilde{r}_1(x), \tilde{r}_2(x), \ldots, \tilde{r}_s(x)]$

$$R_W(x) = R(x)W$$

$$\phi_W(x) = \frac{1}{s}\sum_{j=1}^{s}\|\tilde{r}_j(x)\|^2 = \frac{1}{s}\mathrm{tr}(R_W(x)^T R_W(x))$$

Proposition

If $\mathbb{E}[WW^T] = I$, then

$$\mathbb{E}[\phi_W(x)] = \phi(x), \quad \mathbb{E}[\nabla\phi_W(x)] = \nabla\phi(x)$$

Disadvantage: Conclusion holds only for 2-norm.

Proposition

If $\mathbb{E}[WW^T] = I$, then

$$\mathbb{E}[\phi_W(x)] = \phi(x), \quad \mathbb{E}[\nabla\phi_W(x)] = \nabla\phi(x)$$

Disadvantage: Conclusion holds only for 2-norm.

Sample a small subset $\mathcal{S} \subset \{1, \ldots, m\}$

$$\phi_S(x) = \frac{1}{s} \sum_{i \in \mathcal{S}} \phi_i(x), \nabla \phi_S(x) = \frac{1}{s} \sum_{i \in \mathcal{S}} \nabla \phi_i(x)$$

$$\mathbb{E}[\phi_W(x)] = \phi(x), \quad \mathbb{E}[\nabla \phi_W(x)] = \nabla \phi(x)$$

In each iteration step, cost is about $\frac{s}{m}$ -fraction of the ture cost.

Error Analysis

What we care about are the sample gradient.

$$e := \nabla\phi_S - \nabla\phi$$

$$\mathbb{E}[\|e\|^2] = \mathbb{V}[\|\nabla\phi_S\|]$$

$$\sigma_g := \frac{1}{m-1}\sum_{i=1}^{m}\|\nabla\phi_i - \nabla\phi\|^2$$

$$\mathbb{E}[\|e\|^2] = \frac{1}{s}(1 - \frac{s}{m})\sigma_g \quad \text{(without replacement)}$$

$$\mathbb{E}[\|e\|^2] = \frac{1}{s}\sigma_g \quad \text{(with replacement)}$$