Modified on Wednesday 25th November, 2020, 12:39:4412:39

Nov 25, 2020

- Randomized linear algebra
- Let A and B be large data matrices. We want to approximate AB. A: $m \times n$, B: $n \times p$.
- $A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix}$, $B = \begin{bmatrix} B_1^T & B_2^T & \cdots & B_n^T \end{bmatrix}^T$. A_i: the i-th column of A, B_i: the i-th row of B. Then the product AB satisfies

$$AB = A_1B_1 + A_2B_2 + \cdots + A_nB_n.$$

- Practice with computing mean and variance. X: random variable, the number by throwing a dice. X takes values $1, 2, \ldots, 6$. Probability of $X = i$ is p_i. For a fair dice, $p_i = \frac{1}{6}$.

  Mean, expectation, expected value

$$E[X] = \sum_{i=1}^{6} ip_i = \frac{7}{2}$$

  Variance

$$E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \sum_{i=1}^{6} i^2 p_i - \left(\sum_{i=1}^{6} ip_i\right)^2$$

- Random matrix multiplication with the correct mean AB
- A sampling matrix S of size $n \times s$. Each column of S represent a sample. The size of sampling is s. $C = AS$, $R = S^T B$, $CR = ASS^T B \approx AB$.

  Each column of S has only one nonzero entry, s_{kj}, at a random position $(k, j) = (k(j), j)$.

$$A = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix}, \quad S = \begin{bmatrix} s_{11} & 0 \\ 0 & 0 \\ 0 & s_{32} \end{bmatrix}, \quad AS = \begin{bmatrix} s_{11}A_1 & s_{32}A_3 \end{bmatrix}$$

  There are n unit column vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$. Each column of S is chosen from these n vectors. The probability of choosing $\mathbf{e}_i$ is p_i, with $p_1 + \cdots + p_n = 1$. If A_k and B_k are chosen at the jth trial, multiply A_kB_k by $\frac{1}{sp_k}$, i.e., set $s_{kj} = \frac{1}{\sqrt{sp_k}}$. For a given S, the product CR is

$$CR = \sum_{j=1}^{s} \frac{A_{k(j)}B_{k(j)}}{sp_{k(j)}}$$

  The expectation of CR is AB:

$$E[CR] = E[ASS^T B] = \sum_{j=1}^{s} \sum_{k=1}^{n} p_k \frac{A_kB_k}{sp_k} = AB$$

- Which probability distribution $p_1, \ldots, p_n$ serves the best? We want a probability distribution with small variance.
  - Uniform sampling: $p_k = \frac{1}{n}$
  - Some unequal probabilities depending on the lengths of columns of A and of rows of B.

- Norm-squared sampling minimizes the variance
- Norm-squared sampling uses the following probability

$$p_k = \frac{\|A_k\|\|B_k^T\|}{c}, \quad c = \sum_{k=1}^{n} \|A_k\|\|B_k^T\|.$$

  It turns out that this probability minimizes the variance

$$E\left[\|CR - E[CR]\|_F^2\right] = \frac{1}{s}\left[\sum_{k=1}^{n} \|A_k\|\|B_k^T\|\right]^2 - \frac{1}{s}\|AB\|_F^2 = \frac{1}{s}\left(c^2 - \|AB\|_F^2\right).$$

- How do we get the above result? Let $p_1, \dots, p_n$ be fixed.
  - Mean

$$E[CR] = s\sum_{k=1}^{n} p_k \frac{A_k B_k}{s p_k} = AB$$

  - Variance

$$\begin{aligned} E[\|CR - E[CR]\|_F^2] &= s\sum_{i,l}\left(\sum_{k=1}^{n} p_k \frac{a_{ik}^2 b_{kl}^2}{s^2 p_k^2} - \left(\sum_{k=1}^{n} p_k \frac{a_{ik} b_{kl}}{s p_k}\right)^2\right) \\ &= \sum_{k=1}^{n} \frac{\|A_k\|^2\|B_k^T\|^2}{s p_k} - \frac{1}{s}\|AB\|_F^2 \end{aligned}$$

  - Minimize the above variance with condition $p_1 + \cdots + p_n = 1$.
    Use the Lagrange multiplier method. $f(x_1, \dots, x_n)$ with $g(x_1, \dots, x_n) = 0$. $\mathrm{grad}(f) = \lambda\,\mathrm{grad}(g)$.

$$\left(-\frac{\|A_1\|^2\|B_1^T\|^2}{s p_1^2}, \dots, -\frac{\|A_n\|^2\|B_n^T\|^2}{s p_n^2}\right) = \lambda(1, \dots, 1), \quad p_1 + \cdots + p_n = 1$$

$$p_j = \frac{\|A_j\|\|B_j^T\|}{c}, \quad c = \sum_{j=1}^{n} \|A_j\|\|B_j^T\|,$$

  - Norm-squared sampling uses the optimal probabilities p_j for minimum variance.