Факторизация матриц при помощи нейронных сетей

Аспирант: Гребенюк Александр Андреевич
Научный руководитель: Клышинский Эдуард Станиславович
Год обучения: Первый
Дата: 20.07.2017

Факторизация матриц

- Факторизация (разложение) матрицы – в общем случае, представление исходной матрицы $\mathbf{A}$ в виде произведения

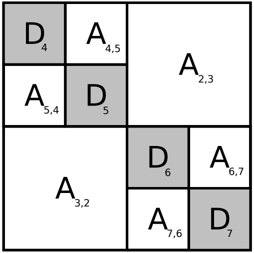
$$\mathbf{A} = \prod_j \mathbf{A}_j.$$

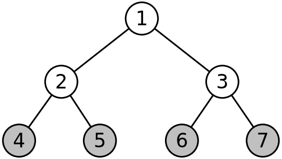
- Примеры:
  - $\mathbf{A} = \mathbf{LDU}$, $\mathbf{L}$ – нижнетреугольная, $\mathbf{D}$ – диагональная, $\mathbf{U}$ – верхнетреугольная матрица;
  - $\mathbf{A} = \mathbf{VDV}^*$, $\mathbf{V}$ – матрица собственных векторов, $\mathbf{D}$ – матрица собственных значений;
- Области применения:
  - Задачи прикадной математики (приведение матриц СЛАУ к определенному виду, предобуславливание)
  - Задачи анализа данных (построение низкоранговых аппроксимаций исходной матрицы, выявление кластеров)

LUP-разложение

- Разложение вида:
$$\mathbf{PA} = \mathbf{LU},$$
$\mathbf{L}$ – нижнетреугольная, $\mathbf{U}$ – верхнетреугольная, $\mathbf{P}$ – матрица перестановок.
- Может применяться для решения СЛАУ следующим образом:
$$\begin{aligned} &\mathbf{A}^* = \mathbf{PA} \\ &\mathbf{A}^*\mathbf{x} = \mathbf{b} \\ &\mathbf{LUx} = \mathbf{b} \\ &\mathbf{Ly} = \mathbf{b} \Longrightarrow \mathbf{y} \sim O(N) \\ &\mathbf{Ux} = \mathbf{y} \Longrightarrow \mathbf{x} \sim O(N) \end{aligned}$$
- Сложность вычисления разложения: $O(\frac{2}{3}N^3)$.

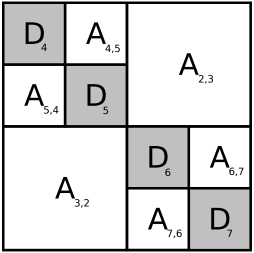
Иерархическое разложение (HSS)

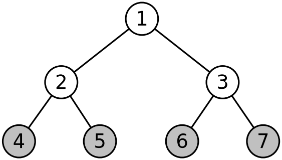




- Матрица представляется в виде бинарного дерева блоков
- Внедиагональные блоки факторизуются в произведение низкоранговых матриц.
- Сложность вычисления разложения: $O(N^2)$

Иерархическое разложение (HSS)





- Ограничения:
  - $rank\left(\mathbf{A}_{i,j}\right) \leq k$;
  - Внедиагональные блоки имеют разложение по базису из k векторов:

$$\mathbf{A}_{i,j} = \mathbf{U}_i^b \mathbf{B}_{i,j} \mathbf{V}_i^{b*},\ \mathbf{B}_{i,j} \in \mathbb{C}_{k \times k}$$

  - $\mathbf{U}_i^b$ и $\mathbf{V}_i^b$ рекуррентно выражаются через матрицы потомков:

$$\mathbf{U}_i^b = \begin{bmatrix} \mathbf{U}_{2i+1}^b & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{2i+2}^b \end{bmatrix} \cdot \mathbf{U}_i$$

Пример использования HSS-разложения

- С использованием HSS-разложения произведение матрицы на вектор $\mathbf{b} = \mathbf{A}\mathbf{x}$ может быть вычислено за линейное время $O(kN)$
- Алгоритм:
  1. Для листьев l вычисляется $\mathbf{x}'_l = \mathbf{V}_l^* \mathbf{x}_l$;
  2. Для узлов v с потомками l_1 и l_2 от листьев к корню вычисляется:
  $$\mathbf{x}'_v = \mathbf{V}_v^* \begin{bmatrix} \mathbf{x}'_{l_1} \\ \mathbf{x}'_{l_2} \end{bmatrix};$$
  3. Для узлов v с потомками l_1 и l_2 от корня к листьям вычисляется:
  $$\begin{bmatrix} \mathbf{b}'_{l_1} \\ \mathbf{b}'_{l_2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{B}_{l_1,l_2} \\ \mathbf{B}_{l_1,l_2} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}'_{l_1} \\ \mathbf{x}'_{l_2} \end{bmatrix} + \mathbf{U}_v \mathbf{b}'_v;$$
  4. Наконец, для листьев l вычисляются части итогового результата: $\mathbf{b}_l = \mathbf{U}_l \mathbf{b}'_l + \mathbf{D}_l \mathbf{x}_l$.

SVD-разложение

- Разложение вида:
$$\mathbf{A} = \mathbf{SVD},$$
$\mathbf{S}, \mathbf{D}$ – унитарные, $\mathbf{V}$ – матрица сингулярных чисел;
- Наибольший интерес с т.з. анализа данных имеет $\mathbf{V}$;
- Путем исключения векторов, отвечающих малым сингулярным числам можно получить наилучшие низкоранговые аппроксимации исходной матрицы;
- Сложность вычисления разложения: $O\left(\min(MN^2, NM^2)\right)$.

Факторизация с использованием нейронных сетей

- Факторизация матриц полезна, но вычислительно затратна (LUP $\sim O(\frac{2}{3}N^3)$, SVD $\sim O\left(\min(MN^2, NM^2)\right)$, HSS $\sim O(N^2)$);
- Часто имеет смысл получения неполных разложений за приемлемое время – для предобуславливания, получения начального приближения при решении СЛАУ, получения части сингулярных чисел и т.д.;
- Для решения такой задачи можно использовать нейронные сети;
- Пример для LU-разложения:

$$loss(\mathbf{M}) = \alpha||\mathbf{U}_M - \mathbf{U}^*||_F + \beta||\mathbf{L}_M - \mathbf{L}^*||_F + \gamma||\mathbf{M} - \mathbf{L}^*\mathbf{U}^*||_F,$$

$$||\mathbf{A}||_F = \sum_{i=0}^{N}\sum_{j=0}^{M} a_{ij}^2, \mathbf{A} \equiv (a_{ij}) \in \mathbb{R}_{N\times M},$$

$||\cdot||_F$ – норма Фробениуса.

Достоинста и недостатки нейросетевого подхода

- Достоинства:
  - Сложность вычисления приближения разложения - номинально $O(N^2)$ и зависит от структуры нейронной сети, но процесс вычисления хорошо параллелизуем, чем не всегда обладают алгоритмы вычисления разложений;
- Недостатки:
  - Размер входных данных ограничен числом узлов во входном слое нейросети;
  - Для снижения затрат по памяти необходимо учитывать разреженность входных данных. Чтобы сделать обучение возможным, необходимо нетривиальное отображение разреженной матрицы во входной и выходной слои нейросети;
  - Приближение будет получено с некоторой точностью, для получения с заданной точностью может потребоваться некоторый итерационный алгоритм вычисления приближения разложения.

Вывод

- Задача факторизации матриц актуальна во множестве областей и вычислительно затратна;
- Несмотря на обилие недостатков, нейросети могут быть применены для решения этой задачи;
- Подобный подход не является широко освещенным в литературе, но есть публикации в которых он показывает приемлемую точность, например в статье M. Roy et al. "Neural Network Matrix Factorization";
- Решение задачи факторизации матриц является подзадачей для решения задачи в теме работы аспиранта.