دانشکده علوم ریاضی
دانشگاه صنعتی شریف

الگوریتم‌های خلاصه‌سازی برای مه‌داده

محمدهادی فروغمنداعرابی
پاییز ۱۳۹۹

ضرب سریع و تقریبی ماتریس

جلسه چهاردهم

نگارنده: عرفان متشرعی

۱ ضرب معمولی ماتریس

فرض کنید ماتریس های A و B به صورت زیر موجود باشند:

$$A = \begin{pmatrix} ----- & a_1^T & ----- \\ ----- & a_2^T & ----- \\ . & . & . \\ . & . & . \\ . & . & . \\ ----- & a_n^T & ----- \end{pmatrix} \qquad B = \begin{pmatrix} ----- & a_1^T & ----- \\ ----- & a_2^T & ----- \\ . & . & . \\ . & . & . \\ . & . & . \\ ----- & a_n^T & ----- \end{pmatrix}$$

$$A \in \mathbb{R}^{n\times p} \qquad B \in \mathbb{R}^{n\times d}$$

هدف به دست آوردن $A^\top B$ است.

اگر برای ماتریس مربعی با روش معمولی ضرب ماتریس عمل کنیم زمان آن از $O(n^3)$ خواهد بود.

در طول تاریخ کار های زیادی برای کم کردن زمان محاسبه دقیق این ضرب انجام شده است که نتایج آن به صورت زیر بوده است.

$$O(n^\omega)$$

۰. $\omega = 3$

۱. $\omega < \log_2 7$ (Strassen)

۲. $\omega < 2/376$ (Coppersmith،Winograd)

۳. $\omega < 2/374$ (Stothevs)

۴. $\omega < 2/3728642$ (Vassilevke-Williams)

۵. $\omega < 2/3728639$ (LeGell)

۲ هدف تقریبی

می خواهیم ماتریس C را به گونه ای بدست آوریم که:

$$C \simeq A^\top B$$

و شرط زیر برای آن برقرار باشد:

$$\left\|A^T B - C\right\|_X < \varepsilon, \text{ probability with } > 1-\delta$$

گزینه های زیر برای نرم آن موجودند:

$$\text{X} : \begin{cases} \|M\|_F = \left(\sum_{i,j} M_{i,j}^2\right)^{\frac{1}{2}} \\ \|M\| = \sup_{\|x\|=1} \left|x^T M x\right| \end{cases}$$

۳ روش اول: نمونه گیری

اگر بتوانیم $A^\top B$ را به صورت جمع n تا عدد بنویسیم و به جای محاسبه همه ی آن ها، از m تا ی آن ها نمونه گیری کنیم، زمان کار بسیار کمتر خواهد شد.

مقدار $A^\top B$ را می توانیم به صورت زیر بنویسیم:

$$A^\top B = \sum_{j=1}^{n} A_{i,j} B_{j,k} = \sum_{j=1}^{n} a_j[i] b_j[k] = \sum_{j=1}^{n} \left(a_j b_j^T\right)_{i,k}$$

و در نتیجه داریم:

$$A^\top B = \sum_{i=1}^{n} a_i b_i^\top$$

حال می توانیم با انتخاب m تا از i های بالا تقریب خوبی از $A^\top B$ به دست آوریم.

اما اگر احتمال انتخاب آن ها برابر باشد، در صورتی که فقط یک یا چند i مهم و بزرگ داشته باشیم، این کار واریانس خوبی نخواهد داشت، پس بهتر است این i ها را با احتمال برابر انتخاب نکنیم.

می توانیم برای رفع این مشکل هر i را با احتمال p_i انتخاب کنیم که متناسب با نرم $a_j b_j^T$ است.

برای این کار می توانیم از ماتریس نمونه گیری Π استفاده کنیم.

$$\Pi = \frac{1}{\sqrt{m}} \begin{pmatrix} 0 & 0 & \dots & \frac{1}{\sqrt{p_{i_1}}} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \frac{1}{\sqrt{p_{i_m}}} & 0 \end{pmatrix}$$

و درنتیجه خواهیم داشت:

$$\Pi A \in \mathbb{R}^{m\times d} \quad \Pi B \in \mathbb{R}^{m\times p}$$

و C خروجی ما می تواند به صورت زیر باشد:

$$C = (\Pi A)^T \Pi B = \frac{1}{m}\sum_{k=1}^{m} \frac{a_{i_k} b_{i_k}^T}{p_{i_k}}$$

دقت کنید که i_k ها به صورت تصادفی و با احتمال p_i انتخاب می شوند.

باید نشان دهیم امید و واریانس C به دست آمده خوب است:

مرحله اول: امید

$$\mathbb{E}\frac{a_{i_k} b_{i_k}^T}{p_{i_k}} = \sum_{j=1}^{n} p_j \frac{a_j b_j^T}{p_j} = A^T B$$

و در نتیجه:

$$\mathbb{E}C = \frac{1}{m}\sum_{k=1}^{m} \mathbb{E}\frac{a_{i_k} b_{i_k}^T}{p_{i_k}} = A^T B$$

مرحله دوم: واریانس هدف:

$$\mathbb{P}\left(\left\|C - A^T B\right\|_F > \varepsilon \|A\|_F \|B\|_F\right) < \eta$$

با استفاده از کران مارکوف داریم:

$$\mathbb{P}\left(\left\|C - A^T B\right\|_F^2 > \varepsilon^2 \|A\|_F^2 \|B\|_F^2\right) < \frac{\mathbb{E}\left\|C - A^T B\right\|_F^2}{\varepsilon^2 \|A\|_F^2 \|B\|_F^2}$$

در نامساوی بالا در سمت راست مخرج کسر ثابت است، کافیست برای صورت کسر کرانی پیدا کنیم.

چون داشتیم:

$$p_i \propto \|a_i\|_2 \|b_i\|_2$$

پس داریم:

$$\frac{\mathbb{E}\left\|C - A^T B\right\|_F^2}{\varepsilon^2 \|A\|_F^2 \|B\|_F^2} < \frac{C}{\varepsilon^2 m}$$

پس کافیست m طوری انتخاب کنیم که:

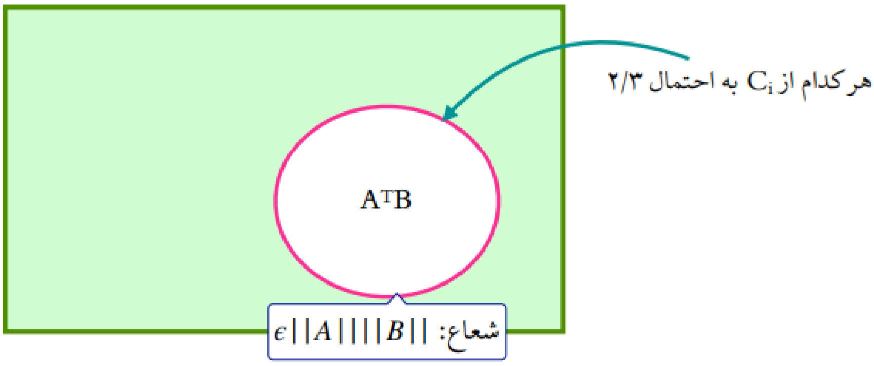
$$m \geq \frac{C}{\varepsilon^2 \eta}$$

برای بهتر شدن زمان اجرا مشابه تکنیکی که در جلسات گذشته استفاده میکردیم عمل میکنیم:

قرار می دهیم $\eta = \frac{1}{3}$ و t بار این کار را تکرار می کنیم.

$$t = \Theta\left(\lg \frac{1}{\delta}\right)$$

حال باید ماتریس C ای را برگزینیم که به اندازه کافی خوب است. در گذشته در مواجه با اعداد در این تکنیک، میانه را انتخاب میکردیم، اما چالش اصلی ما در اینجا این است که خروجی این t بار تکرار، هرکدام یک ماتریس است. خوشبختانه می توانیم از ایده زیر استفاده کنیم:



فضای ماتریس ها را به شکل بالا در نظر بگیرید. اگر یک کره به شعاع گفته شده و به مرکز $A^\top B$ بزنیم، هر کدام از C_i های به دست آمده از هر بار اجرا به احتمال $\frac{۲}{۳}$ در این کره قرار می گیرد. و به آسانی میتوان نشان داد که به احتمال خوبی حداقل $\frac{t}{۲}$ از ماتریس ها در آن قرار می گیرند. پس اگر از بین آن ها ماتریسی را بیابیم که با حداقل $\frac{۲}{۳}$ ماتریس دیگر فاصله ی کمی دارد کار تمام است، که به طور دقیق تر می توانیم به صورت زیر بیان کنیم:

$$S_i = \left|\{j : \|C_i - C_j\|_F \leq ۲\varepsilon\|A\|_F\|B\|_F\}\right|$$

C_i خوب است اگر:

$$S_i \geq \frac{t}{۲}$$

۱.۳ تحلیل زمان اجرا

تولید ΠA و ΠB: کمتر از خود ماتریس ها هر ضرب ΠA و ΠB:

$$\mathrm{O}\left(\mathrm{pd}\epsilon^{-۲}\right) = \mathrm{O}(\mathrm{pmd})$$

اجرای مستقل برای t تا ΠA و ΠB: همه ی ضرب ها:

$$\mathrm{O}\left(\mathrm{pd}\log ۱/\delta\epsilon^{-۲}\right)$$

انتخاب C_i خوب:

$$\mathrm{O}\left(\mathrm{pd}\log^{۲} ۱/\delta\right)$$

در نتیجه در مجموع:

$$\mathrm{O}\left(\mathrm{pd}\log^{۲} ۱/\delta\right)$$

۴ روش دوم: خلاصه‌سازی خطی غافل

ایده: این روش بر خلاف روش اول، Π را مستقل از A و B به دست می آورد.
به طور کامل در جلسه ی بعد گفته خواهد شد.

۵ منابع

مباحث جلسه چهاردهم