دانشکده علوم ریاضی
دانشگاه صنعتی شریف

الگوریتم‌های خلاصه‌سازی برای مه‌داده

محمدهادی فروغمنداعرابی
پاییز ۱۳۹۹

مروری بر تجزیه SVD و نشاندن زیرفضا، حل مسئله‌ی رگرسیون

جلسه شانزدهم

نگارنده: مائده حشمتی

۱ مروری بر مباحث گذشته

در جلسه گذشته مروری بر خلاصه‌سازی خطی غافل داشتیم و با نشاندن زیرفضا جلسه را به پایان رساندیم. در این جلسه با کمک ابزارهایی همچون نشاندن زیرفضا و تجزیه‌ی SVD به حل مسئله رگرسیون می‌پردازیم.

۲ تجزیه SVD :

قضیه ۱: هر ماتریس حقیقی $A \in \mathbb{R}^{n\times d}$ با رتبه‌ی r می‌تواند به شکل زیر تجزیه شود:

$$A = U\Sigma V^T$$

به طوریکه $U \in \mathbb{R}^n \times r, V \in \mathbb{R}^{d\times r}, U^TU = I, V^TV = I, \Sigma = diag(\sigma_1, \sigma_2, ..., \sigma_r), \sigma_i > 0$
و σ_i ها را مقادیر تکین ماتریس A می‌نامند و برابر رادیکال مقادیر ویژه‌ی A^TA می‌باشند.
براساس این تجزیه، برای هر ماتریس حقیقی، شبه وارون نیز تعریف می‌شود.

تعریف ۱: برای یک ماتریس حقیقی داده شده با تجزیه‌ی $M = U\Sigma V^T$ ، شبه وارون این ماتریس (pseudoinverse) برابر است با: $M^+ = V\Sigma^{-1}U^T$ (مقادیر تکین در حالت معکوس در Σ^{-1} قرار می‌گیرند: $\Sigma^{-1} = diag(1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_r)$)

ادعا ۱: ماتریس $A(A^TA)^+A^T$ تصویر در زیرفضای ستون‌های A می‌باشد.
اثبات ادعا ۱: با کمک تجزیه‌ی SVD اثبات می‌کنیم:

$$A = U\Sigma V^T A^TA = V\Sigma U^T U\Sigma V^T = V\Sigma^۲ V^T (A^TA)^+ = V\Sigma^{-۲}V^T$$

در نهایت خواهیم داشت:

$$A(A^TA)^+A^T = U\Sigma V^T(V\Sigma^{-۲}V^T)V\Sigma U^T = UU^T$$

که یعنی ضرب داخلی در پایه‌ها و نشان‌دهنده‌ی تصویر از A می‌باشد و حکم ثابت می‌شود.

۳ نشاندن زیرفضا :

تعریف ۲: زیرفضای خطی $E \subset \mathbb{R}^n$ را در نظر بگیرید. ماتریس $\Pi \in \mathbb{R}^{m\times n}$ یک نشاننده برای این زیرفضاست ($\varepsilon-subspace\ embedding$) اگر داشته باشیم:

$$\forall x \in E : (۱-\varepsilon)\|x\|_۲^۲ \le \|\Pi x\|_۲^۲ \le (۱+\varepsilon)\|x\|_۲^۲$$

یعنی ماتریسی که تقریبا طول بردارهای درون این زیرفضا را ثابت نگه‌می‌دارد.

تعریف ۳: یک نشاننده‌ی زیرفضای غافل $(\varepsilon, d, \delta) - oblivious\ subspace\ embedding\ (OSE)$ ، یک توزیعی D روی ماتریس‌های $\Pi \in \mathbb{R}^{m\times n}$ است به طوریکه به ازای هر ماتریس $U \in \mathbb{R}^{n\times d}$ با ستون‌های متعامد داریم:

$$\mathbb{P}_{\Pi\sim D}(\|(\Pi U)^T(\Pi U) - I\| > \varepsilon) < \delta$$

۴ رگرسیون :

هدف این است که با توجه به داده‌هایی که تا به حال داشتیم و قضایا، یک مسئله‌ی رگرسیون را در زمان کمتری حل کنیم. ابتدا به تعریف مسئله‌ی رگرسیون می‌پردازیم.

تعریف ۴: در یک مسئله‌ی رگرسیون، یک ماتریس ویژگی‌ها داریم که سطرها، نمونه‌های ما هستند و تعداد آنها بسیار زیاد است و ستون‌های آن، ویژگی‌های مورد نظر ما می‌باشد که تعدادش از سطرها بسیار کمتر است. هدف یافتن یک ماتریسی است که در صورت ضرب شدن در ماتریس ما، بتوانیم جواب نهایی را پیشبینی کنیم. یعنی با یک ترکیب خطی خوب از سطرهای ماتریس داده‌هایمان، به جواب نهایی با صحت بالا برسیم:

$$X\beta \simeq y$$

X همان ماتریس ویژگی‌ها و داده‌های اولیه‌ی ما، y هدف پیش بینی و β مجهول ما است. $(X \in \mathbb{R}^n \times d\ :\ n >> d)$ پیشنهاد اولیه برای حل این مسئله در زمان کمتر، استفاده از کمترین مربعات می‌باشد.

$$\beta^{LS} = argmin_\beta \|X\beta - y\|_۲, \quad X \in \mathbb{R}^{n\times d},\ y \in \mathbb{R}^n$$

می‌دانیم هر بردار، نمایشی از برداری درون زیرفضای ساخته شده از ستون‌های X و برداری درون زیرفضای عمود برآن دارد، ($y = y^\perp + y^\|$) ، خواهیم داشت:

$$\|X\beta - y\|_۲^۲ = \|X\beta - y^\perp - y^\|\|_۲^۲$$
$$= \|X\beta - y^\|\|_۲^۲ - ۲\langle X\beta - y^\|, y^\perp\rangle + \|y^\perp\|_۲^۲$$

جمله دوم به وضوح صفر است و چون $y^\perp$ معین می‌باشد، آنچه دست ماست تنها جمله‌ی اول یعنی $\|X\beta - y^\|\|_۲^۲$ می‌باشد. می‌توانیم β را به نحوی اتخاذ کنیم که داشته باشیم: $X\beta = y^\|$ و برای تصویر y داریم: $X(X^TX)^+X^Ty$ در نهایت برای β^{LS} داریم: $\beta^{LS} = (X^TX)^+X^Ty$.

در این حالت زمان اجرای ما بسیار زیاد بود و در ادامه تلاش خواهیم کرد که تعداد سطرها را به نحوی کاهش دهیم. در این صورت به یک رگرسیون جدید نیاز داریم به طوری که به جای X از ΠX و به جای y از Πy استفاده کنیم. ماتریس Π باید به نحوی باشد که تعداد صفرهایش زیاد و به تعداد کمی ۱ یا ۱ـ داشته باشد تا زمان اجرا را کاهش دهد و نیز جواب رگرسیون اصلی را زیاد تغییر ندهد. به طور کلی چنین ایده‌ای داریم:

$$\hat{\beta}^{LS} = argmin\|\Pi X\beta - \Pi y\|_2^2$$

و تصویر y برابر خواهد بود با: $\Pi X(\Pi X^T \Pi X)^{+}\Pi X^T \Pi y$

لم ۱ : زیرفضای خطی $E := span(cols(X), y)$ را در نظر بگیرید. ماتریس Π یک نشاننده ی این زیرفضاست. در اینصورت خواهیم داشت:

$$\|X\hat{\beta}^{LS} - y\|_2^2 \leq \frac{1+\varepsilon}{1-\varepsilon}.\|X\beta^{LS} - y\|_2^2$$

اثبات لم ۱ : با توجه به تعریف نشاندن و تعریف $\hat{\beta}$ خواهیم داشت:

$$(1-\varepsilon)\|X\hat{\beta}^{LS} - y\|_2^2 \leq \|\Pi X\hat{\beta}^{LS} - \Pi y\|_2^2 \leq \|\Pi X\beta^{LS} - \Pi y\|_2^2 \leq (1+\varepsilon)\|X\beta^{LS} - y\|_2^2$$

پس حالا نیاز داریم تا یک Π بیابیم تا در شروط ما صدق کند و بتواند جوابی برای مسئله رگرسیون بدهد.

قضیه ۲ : برای ماتریس $X \in \mathbb{R}^{n\times d}$ و بردار $y \in \mathbb{R}^n$ در زمان $O(nnz(X) + n) + poly(d/\varepsilon)$ میتوان با احتمال ۰٫۹ $\hat{\beta}^{LS}$ را به گونه‌ای یافت که در شرایط زیر صدق کند:

$$\|X\hat{\beta}^{LS} - y\|_2^2 \leq \frac{1+\varepsilon}{1-\varepsilon}.\|X\beta^{LS} - y\|_2^2$$

یکی از انواع ماتریس‌های مطلوب برای ما، ماتریسی است که خاصیت OSE را دارد و در شروط زیر صدق می‌کند:

$$\mathbb{P}_{\Pi\sim D}(\|(\Pi U)^T(\Pi U) - I\| > \varepsilon) < \delta$$

یکی از این دسته، ماتریس‌های $countsketch$ می‌باشد که در هر ستون با یک تابع هش، مکان تنها درایه‌ی ناصفر مشخص می‌شود و با احتمال ۱/۲ برابر ۱+ یا ۱ـ خواهد شد. این ماتریس با تعداد سطرهای $m = d^2/\varepsilon^2$ خاصیت OSE را دارد و در شروط مورد نظر ما صدق می‌کند.

۵ احساس فشردگی :

صورت کلی مسئله‌ی $Compressed\ Sensing$ یا احساس فشردگی، بدین صورت می‌باشد که یک بردار بزرگ و k تنک داریم و چون برای ذخیره‌سازی این بردار، به فضای زیادی نیاز داریم، میخواهیم به جای این بردار، Πx را ذخیره کنیم تا فضای کمتری اشغال کند و در صورت نیاز به x، عملیات بازیابی را به نحوی انجام دهیم که با تقریب خوبی به x اولیه برسیم. ماتریس Π باید به نحوی باشد که تعداد ستون‌هایش برابر با تعداد سطرهای x باشد، اما تعداد سطرهایش به قدر کافی کوچک باشد تا در نهایت ضرب این دو، به برداری با طول کمتر رسیده تا فضای کمی برای ذخیره‌سازی اشغال کند.

$$x \longrightarrow y = \Pi x$$

و برای بازیابی x از روی y و Π :

$$\begin{aligned} &min\ \|z\|_1 \\ &s.t.\quad \Pi z = y \end{aligned}$$

در واقع با این الگوریتم، به دنبال تنک ترین z می‌گردیم که حاصلضرب Πz برابر با مقدار ذخیره شده یعنی y شود.
علت قرار دادن نرم یک در تابع هدف: قرار دادن نرم صفر که دقیقاً تعداد درایه‌های ناصفر بردار را می‌دهد در اینجا موجب سخت شدن مسئله خواهد شد و مسئله به $NP-Complete$ تبدیل می‌شود. بقیه‌ی نرم‌ها مانند دو از نظر هندسی برای رسیدن به جواب، مطلوب نیستند و ساده ترین و بهترین نرم برای یافتن تنک ترین بردار، نرم یک می‌شود.

این الگوریتم برای حالتی که x حقیقی باشد و به جای دقیقاً تنک، تقریبا تنک نیر باشد کار می‌کند. در واقع با توجه به خواسته‌ی ما و ورودی مسئله، این الگوریتم می‌تواند برای ما مطلوب باشد. به طور مثال، می‌توانیم با این روش، k تا بزرگترین درایه‌های x را بازیابی کنیم و می‌دانیم این روش، تقریباً جواب خوبی به ما خواهد داد.

تعریف ۵: ماتریس $\Pi \in \mathbb{R}^{m\times n}$ دارای خاصیت $(\varepsilon, k) - restricted\ isometry\ property$ و یا به طور خلاصه RIP است، اگر به ازای تمام بردارهای k تنک با طول واحد در شرط زیر صدق کند:

$$1 - \delta \leq \|\Pi x\|_2^2 \leq 1 + \delta$$

یعنی ماتریسی که برای بردارهای k تنک، طول بردار را تقریباً حفظ کند. به صورت دیگرف همانطور که در جلسات گذشته نیز گفتیم یعنی داشته باشیم:

$$sup_{T \subset [n] | |T| = k} \quad \|I_k - (\Pi^{(T)})^* \Pi^{(T)}\| < \delta$$

تعریف ۶: یک ماتریس A $m \times n$ دارای خاصیت $null - space\ property\ of\ order\ k$ با ثابت C می‌باشد اگر به ازای هر $\eta \in \Re^n$ که $A\eta = 0$ و به ازای هر مجموعه k عضوی $T \subset \{1, 2, ..., n\}$ داشته باشیم:

$$\|\eta\|_1 \leq C \|\eta_{-T}\|_1$$

$-T$ به معنای مجموعه‌ی مکمل T است.
این ویژگی بدین معنی است که به ازای بردار η درایه‌هایش بسیار بزرگ نباشند و بتوان k تا از بزرگترین درایه‌هایش را با بقیه‌ی درایه‌ها کران کرد و طبق تعریف، زیرفضای پوچ A دارای این خاصیت می‌باشد.

لم ۲: اگر ماتریس A دارای خاصیت RIP از $order$ برابر با $k(2 + c)$ و ثابت برابر با $\delta, c > 1$ باشد، آنگاه A دارای خاصیت $null - space\ property$ از $order$ برابر با $2k$ و ثابت $C = 1 + \sqrt{2/c}(1 + \delta)(1 - \delta)$ خواهد بود.
یعنی خاصیت RIP خاصیت $null - space$ را نتیجه می‌دهد.

اثبات لم بالا در جلسه آینده گفته خواهد شد.