مقدمه‌ای بر یادگیری ماشین

پاییز ۱۴۰۱

اساتید:علی شریفی، بهروز آذرخلیلی

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی کامپیوتر

پاسخنامه کوییز اول

تاریخ برگزاری: ۱ آذر

۱:

- (L1-norm)

$$A = \begin{bmatrix} ۲ & -۸ \\ ۳ & ۱ \end{bmatrix}$$

$$\| A \|_۱ = max(۲+۳, ۸+۱) = ۹$$

- (infinity-norm)

$$A = \begin{bmatrix} ۳ & ۶ & -۱ \\ ۳ & ۱ & ۰ \\ ۲ & ۴ & -۷ \end{bmatrix}$$

$$\| A \|_\infty = max(۳+۶+۱, ۳+۱+۰, ۲+۴+۷) = ۱۳$$

۲:

الف) نادرست

مقدار پارامتر C مثل یک پیچ تنظیم کار می کند و با تغییر مقدار آن تعداد داده هایی که می توانند داخل حاشیه قرار بگیرند را تنظیم می کنیم. با کاهش مقدار C اجازه می‌دهیم تعداد ξ_iها افزایش یابد و در نتیجه حاشیه بزرگتر شده و تعداد داده های بیشتری داخل حاشیه قرار گیرد و لذا خطای آموزش افرایش می یابد.

ب) درست

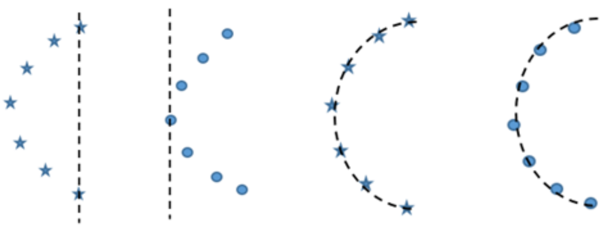
با افزایش مقدار k داده های بیشتری برای آموزش مدل باقی می ماند و در نتیجه مدل بهتر آموزش داده می شود و خطای تخمین زده شده کاهش می یابد.

ج) درست

الگوریتم K-means به صورت تضمینی به یک مینیمم محلی (نه لزوماً سراسری) همگرا می شود و به انتخاب مقدار k و محل مراکز اولیه حساس است زیرا این دو مقدار نقش تعیین کننده در همگرا شدن به یک نقطه بهینه سراسری دارند.

د) نادرست

وقتی از کرنل خطی استفاده می کنیم بردارهای پشتیبان، داده های نشان داده شده در شکل زیر هستند:



اما اگر از کرنل استفاده کنیم همه ی داده ها بردار پشتیبان خواهند بود.(شکل بالا سمت راست)

ه) نادرست.

تعداد داده ها بسیار کم و تعداد ویژگی ها بسیار زیاد است، لذا مایلیم برای پیچیده نشدن مدل تعداد زیادی از ویژگی ها را حذف کنیم. روش L1-Regularization باعث صفر شدن تعداد بیشتری از ویژگی ها نسبت به روش L2-Regularization می شود پس انتخاب L1-Regularization مناسبتر است.

:۳

$$\bar{\mathbf{x}} = \frac{۱}{N}\mathbf{X}^\top ۱ = \frac{۱}{۳}\begin{bmatrix} -۲ & -۱ & -۱ \\ ۱ & ۳ & -۲ \end{bmatrix}\begin{bmatrix} ۱ \\ ۱ \\ ۱ \end{bmatrix} = \begin{bmatrix} \frac{-۴}{۳} \\ \frac{۲}{۳} \end{bmatrix} \tag{۱}$$

$$\mathbf{T} = \mathbf{X} - ۱\bar{\mathbf{x}}^\top = \begin{bmatrix} -۲ & ۱ \\ -۱ & ۳ \\ -۱ & -۲ \end{bmatrix} - \begin{bmatrix} \frac{-۴}{۳} & \frac{۲}{۳} \\ \frac{-۴}{۳} & \frac{۲}{۳} \\ \frac{-۴}{۳} & \frac{۲}{۳} \end{bmatrix} = \begin{bmatrix} \frac{-۲}{۳} & \frac{۱}{۳} \\ \frac{۱}{۳} & \frac{۷}{۳} \\ \frac{۱}{۳} & \frac{-۸}{۳} \end{bmatrix} \tag{۲}$$

$$\mathbf{S} = \frac{۱}{N-۱}\mathbf{T}^\top\mathbf{T} = \frac{۱}{۲}\begin{bmatrix} \frac{-۲}{۳} & \frac{۱}{۳} & \frac{۱}{۳} \\ \frac{۱}{۳} & \frac{۷}{۳} & \frac{-۸}{۳} \end{bmatrix}\begin{bmatrix} \frac{-۲}{۳} & \frac{۱}{۳} \\ \frac{۱}{۳} & \frac{۷}{۳} \\ \frac{۱}{۳} & \frac{-۸}{۳} \end{bmatrix} = \begin{bmatrix} \frac{۳}{۹} & \frac{۱}{۱۸} \\ \frac{۱}{۱۸} & \frac{۵۷}{۹} \end{bmatrix} \tag{۳}$$

$$\mathbf{Su} = \lambda\mathbf{u} \implies (\mathbf{S} - \lambda\mathbf{I})\mathbf{u} = ۰ \implies det(\mathbf{S} - \lambda\mathbf{I}) = ۰ \implies det(\begin{bmatrix} \frac{۳}{۹} - \lambda & \frac{۱}{۱۸} \\ \frac{۱}{۱۸} & \frac{۵۷}{۹} - \lambda \end{bmatrix}) = ۰ \tag{۴}$$

$$\implies \lambda^۲ - \frac{۶۰}{۹}\lambda + \frac{۶۸۳}{۳۲۴} = ۰ \implies \begin{cases} \lambda_۱ = \frac{۱۰}{۳} + \frac{\sqrt{۲۹۱۷}}{۱۸} \\ \lambda_۲ = \frac{۱۰}{۳} - \frac{\sqrt{۲۹۱۷}}{۱۸} \end{cases} \tag{۵}$$

(۶) از آنجا که $\lambda_۱ > \lambda_۲$ برای یافتن اولین مؤلفه اصلی می‌بایست بردار ویژه متناظر با $\lambda_۱$ را بیابیم $\implies$

$$\xRightarrow{\lambda = \frac{۱۰}{۳} + \frac{\sqrt{۲۹۱۷}}{۱۸}} \begin{bmatrix} \frac{۳}{۹} - (\frac{۱۰}{۳} + \frac{\sqrt{۲۹۱۷}}{۱۸}) & \frac{۱}{۱۸} \\ \frac{۱}{۱۸} & \frac{۵۷}{۹} - (\frac{۱۰}{۳} + \frac{\sqrt{۲۹۱۷}}{۱۸}) \end{bmatrix}\begin{bmatrix} u_۱ \\ u_۲ \end{bmatrix} = ۰ \tag{۷}$$

$$\mathbf{u}_۱ = \begin{bmatrix} \frac{\sqrt{۲۹۱۷} - ۵۴}{\sqrt{۵۸۳۴ + ۱۰۸\sqrt{۲۹۱۷}}} \\ \frac{۱}{\sqrt{۵۸۳۴ + ۱۰۸\sqrt{۲۹۱۷}}} \end{bmatrix} \tag{۸}$$

$$\mathbf{X}^* = \mathbf{Tu}_۱ = \begin{bmatrix} \frac{۱۰۹ - ۲\sqrt{۲۹۱۷}}{۳\sqrt{۵۸۳۴ + ۳۲۴\sqrt{۲۹۱۷}}} \\ \frac{\sqrt{۲۹۱۷} - ۴۷}{۳\sqrt{۵۸۳۴ + ۳۲۴\sqrt{۲۹۱۷}}} \\ \frac{\sqrt{۲۹۱۷} - ۶۲}{۳\sqrt{۵۸۳۴ + ۳۲۴\sqrt{۲۹۱۷}}} \end{bmatrix} \tag{۹}$$

:۴

از آنجایی که ماتریس A متقارن و مثبت نیمه معین است، می‌توان آن را به صورت زیر، با استفاده از تجزیه به مقادیر ویژه، قطری‌سازی نمود که مقادیر ویژه آن نیز نامنفی هستند:

$$A = Q\Lambda Q^T = Q\Lambda^{\frac{۱}{۲}}\Lambda^{\frac{۱}{۲}}Q^T = (Q\Lambda^{\frac{۱}{۲}})(Q\Lambda^{\frac{۱}{۲}})^T = MM^T$$

که $M = Q\Lambda^{\frac{۱}{۲}}$ می‌باشد. پس، تابع کرنل را می‌توان به صورت زیر بازنویسی نمود:

$$k(x,y) = x^T MM^T y = (M^T x)^T (M^T y) = < M^T x, M^T y >$$

پس نگاشت متناظر این تابع کرنل به صورت $\phi(x) = M^T x$ باشد و در نتیجه، از آنجایی که توانستیم تابع کرنل را به صورت ضرب داخلی در فضای جدید بنویسیم، k یک کرنل معتبر است.