به نام خداوند بخشنده مهربان

جبر خطی

دانشکده مهندسی کامپیوتر

حمیدرضا ربیعی، مریم رمضانی
بهار ۱۴۰۲

مقادیر تکین، تخمین، قطری‌سازی و مشتق

تمرین پنجم و ششم

تاریخ انتشار: ۱۷ خرداد ۱۴۰۲

۱. پرسش‌های خود در مورد این تمرین را در سامانه کوئرا مطرح کنید.

۲. **سیاست ارسال با تاخیر پاسخ:** شما در مجموع در طول نیم‌سال می‌توانید تا سقف ۱۵ روز تمارین خود را با تاخیر ارسال کنید. پس از آن به ازای هر ساعت تاخیر ۵.۰ درصد نمره کسر خواهد شد. مقیاس محاسبه تاخیر ساعت است و تاخیر شما رو به بالا گرد خواهد شد تا به مقیاس ساعت درآید. به عنوان مثال ۲ ساعت و ۱۰ دقیقه تاخیر معادل ۳ ساعت در نظر گرفته خواهد شد. تمارین تئوری و عملی در محاسبه تاخیرها دو تمرین جداگانه در نظر گرفته خواهند شد. یعنی ۲ روز تاخیر بر تمرین تئوری و ۱ روز تاخیر بر تمرین عملی در مجموع برابر ۳ روز در نظر گرفته خواهد شد. در نهایت اگر بیش از ۱۵ روز تاخیر داشته باشید، تاخیر مجاز شما به نحو بهینه‌ای توزیع خواهد شد که کمترین میزان نمره را از دست بدهید.

۳. **سیاست مشارکت دانشجویان در حل تمارین:** دانشجویان می‌توانند در حل تمارین برای رفع ابهام و یا بدست آوردن ایده کلی با یک دیگر مشارکت و هم‌فکری کنند. این کار مورد تایید و تشویق تیم ارائه درس می‌باشد چرا که هم‌فکری و کار گروهی می‌تواند موجب تقویت یادگیری شود. اما بدست آوردن جزئیات راه‌حل و نگارش پاسخ باید تماما منحصرا توسط خود دانشجو انجام شود. بیشینه اندازه مجاز گروهی که در آن می‌توانید برای حل تمرین هم‌فکری کنید برابر ۴ نفر است. رعایت آداب‌نامه‌ی انجام تمرین‌های درسی الزامی است. در صورت مشاهده تخلف مطابق قوانین دانشکده و دانشگاه برخورد خواهد شد.

سوالات تئوری (۲۸۰ نمره)

تاریخ تحویل: ۶ تیر ۱۴۰۲

پرسش ۱ (۴۰ نمره) مساله بهینه‌سازی کمترین مربعات با عبارت منظم‌سازی نرم دوم را در نظر بگیرید.

$$w^* = \texttt{argmin}_w f(w)$$
$$f(w) = \frac{۱}{۲n}\|X^T w - y\|_۲^۲ + \frac{\lambda}{۲}\|w\|_۲^۲$$

برای حل مساله فوق از روش گام‌به‌گام کاهش گرادیانی استفاده می‌کنیم. به این صورت که در هر مرحله در خلاف جهت گرادیان حرکت می‌کنیم تا به یک کمینه محلی برای مساله بهینه‌سازی برسیم. مقدار اولیه را به صورت تصادفی انتخاب می‌کنیم.

$$w^{t+۱} = w^t - \alpha \nabla f(w^t)$$

مقدار α یک مقدار دل‌خواه است. مقدار آن را به صورت زیر در نظر می‌گیریم.

$$\alpha = \frac{۱}{\sigma_{max}(A)}$$

که در آن $\sigma_{max}(A)$ بزرگ‌ترین مقدار تکین ماتریس A است که به صورت زیر تعریف می‌شود.

$$A = \frac{۱}{n} X X^T + \lambda I$$

(آ) (۱۰ نمره) ثابت کنید

$$\nabla f(w) = Aw - \frac{۱}{n} Xy = A(w - w^*)$$

(ب) (۱۰ نمره) ثابت کنید که ماتریس A نیمه مثبت معین است.

(ج) (۱۰ نمره) ثابت کنید

$$\sigma_{max}(I - \alpha A) = ۱ - \alpha \sigma_{min}(A) = ۱ - \frac{\sigma_{min}(A)}{\sigma_{max}(A)}$$

(د) (۱۰ نمره) ثابت کنید

$$\|w^{t+۱} - w^*\| \leq (۱ - \frac{\sigma_{min}(A)}{\sigma_{max}(A)})\|w^t - w^*\|$$

پرسش ۲ (۴۰ نمره) ماتریس

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

را در نظر بگیرید که در آن A و B ماتریس‌های متقارن هستند و ماتریس A غیرتکین شبه مثبت معین است. ثابت کنید که ماتریس M شبه‌مثبت‌معین است اگر و تنها اگر ماتریس $C - B^T A^{-1} B$ شبه‌مثبت‌معین باشد.

پرسش ۳ (۴۰ نمره) فرض کنید که می‌خواهیم در مساله کمترین مربعات، برای هر داده یک وزن به عنوان ضریب اهمیت در نظر بگیریم و تابع زیر را کمینه کنیم.

$$\sum_{i=1}^{n} c_i^2 (w.x_i - y_i)^2$$

فرم بسته پاسخ این مساله را بدست آورید.

پرسش ۴ (۴۰ نمره) ماتریس A را یک ماتریس متقارن $n \times n$ در نظر می‌گیریم. در این صورت ثابت کنید که سه شرط زیر معادل یک‌دیگر هستند.

(آ) A یک ماتریس مثبت نیمه معین می‌باشد.

(ب) تمام مقادیر ویژه ماتریس A مقداری نامنفی دارند.

(ج) ماتریس حقیقی $n \times n$ مانند B یافت می‌شود به گونه‌ای که $A = BB^T$.

پرسش ۵ (۴۰ نمره) نرم فروبینیوس یک ماتریس $A \in \mathbb{R}^{n \times n}$ به صورت $\|A\|_F = \sqrt{\mathbf{Tr}(A^T A)}$ تعریف می‌شود.

(آ) (۵ نمره) نشان دهید

$$\|A\|_F = \left(\sum_{i,j} |A_{ij}|^2 \right)^{\frac{1}{2}}$$

به عبارتی نرم فروبینیوس، نرم اقلیدسی فرم برداری ماتریس A است.

(ب) (۵ نمره) نشان دهید اگر U و V یکه متعامد باشند، آنگاه

$$\|UA\|_F = \|AV\|_F = \|A\|_F$$

(ج) (۱۵ نمره) نشان دهید

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$$

که $\sigma_1, \sigma_2, \cdots, \sigma_r$ مقادیر تکین[1] A هستند.

(د) (۱۵ نمره) سپس نشان دهید

$$\sigma_{\max}(A) \leq \|A\|_F \leq \sqrt{r}\sigma_{\max}(A)$$

پرسش ۶ (۴۰ نمره) اگر A و X ماتریس‌هایی مربعی باشند، مشتق زیر را محاسبه کنید.

$$\frac{\partial \left(det(A) A^T X A \right)}{\partial A}$$

پرسش ۷ (۴۰ نمره) با توجه به عملگرهای ماتریسی به سوالات زیر پاسخ دهید.

الف) به ازای $x \in \mathbb{R}^n$ و $f : \mathbb{R} \to \mathbb{R}$ را طوری تعریف میکنیم که به ازای هر عضو آن داریم: $y_i = f(x_i)$. $\frac{d\mathbf{y}}{dx}$ را محاسبه کنید.

ب) به ازای $x, y \in \mathbb{R}^n$ تعریف میکنیم $z = x \odot y$، $\frac{dz}{dx}$ را محاسبه کنید.

پ) به ازای $x, y \in \mathbb{R}^n$ $\mathbf{z}$ را طوری تعریف میکنیم که به ازای هر عضو آن داریم: $z_i = f(x_i, y_i)$. $\frac{d\mathbf{z}}{dx}$ را محاسبه کنید.

سوالات عملی (۳۲ نمره)

تاریخ تحویل: ۶ تیر ۱۴۰۲

پرسش ۱ (۳۲ نمره) در کاربردهای یادگیری ماشین، یکی از مهم‌ترین مراحل یادگیری بازنمایی[2] مناسب برای داده است. در بسیاری از موارد داده در وضعیتی قرار دارد که برای ماشین قابل تحلیل نیست. به عنوان مثال کار با داده متنی برای ماشین بسیار سخت است. در مواردی دیگر ممکن است داده شکل عددی داشته باشد ولی حجم آن بزرگ باشد و غیر قابل تحلیل باشد. به عنوان مثال ماتریس مجاورت یک گراف فرم عددی دارد ولی تحلیل از جهت بزرگی بازنمایی مناسبی برای هر راس یا کل گراف بدست نمی‌دهد.

یکی از روش‌هایی که به طور مرسوم در یادگیری ماشین برای یادگیری بازنمایی استفاده می‌شود، استفاده از تجزیه ماتریسی است. برای این منظور روش‌های مختلفی از جمله تجزیه نامنفی[3] و یا تجزیه مقدار تکین[4] وجود دارد. در ادامه این سوال می‌خواهیم استفاده از تجزیه مقدار تکین برای یادگیری بازنمایی مناسب را بیان کنیم و در ادامه آن را برای حل یک مساله ساده به کار ببندیم.

---

[1] Singular values
[2] Representation
[3] Non-negative matrix factorization
[4] Singular Value Decomposition

فرض کنید که n داده هر یک با m ویژگی در اختیار داریم. این دادگان را می‌توانیم در فرم یک ماتریس $m \times n$ نمایش دهیم که هر سطر مربوط به یک ویژگی و هر ستون مربوط به یک داده باشد. فرض کنید که این ماتریس d مقدار تکین داشته باشد. بنابرین تجزیه مقدار تکین آن به فرم $U\Sigma V^T$ خواهد بود که در آن U یک ماتریس $m \times d$ و Σ یک ماتریس $d \times d$ قطری و V یک ماتریس $n \times d$ خواهد بود. ماتریس U برای هر یک از دادگان یک بازنمایی با بعد d معرفی می‌کنیم. اگر مقادیر تکین به ترتیب نزولی رو قطر ماتریس Σ چیده شده باشند، می‌توانیم بازنمایی را با انتخاب درایه‌های متناظر با d' مقدار تکین بزرگ‌تر، به بعد d' کاهش دهیم. بنابرین در عین حال می‌توانیم کاهش ابعاد نیز انجام دهیم.

در این مساله می‌خواهیم از این روش یادگیری بازنمایی برای تحلیل دادگان متنی استفاده کنیم. ماتریس هم‌آیی عبارت‌ـسند را در نظر بگیرید که هر درایه آن نشان‌دهنده تکرار یک عبارت (کلمه) در یک سند خواهد بود. با استفاده از تجزیه مقدار تکین برای هر جمله می‌توانیم یک بازنمایی مناسب پیدا کنیم. حال می‌توانیم با استفاده از فاصله کوسینوسی شباهت هر یک از جملات به یک دیگر را پیدا کنیم.

ورودی

در سطر اول ورودی دو عدد a و b و d داده می‌شود که به ترتیب اندازه تعداد اسناد در پایگاه داده و تعداد پرسمان‌ها است. در $a + b$ خط بعدی جملاتی به عنوان اسناد مورد نظر داده می‌شوند. d نیز بعد بازنمایی است که باید یادگرفته شود.

خروجی

در b خط باید شماره شبیه‌ترین جمله به جمله i ام پرسمان‌ها چاپ شود. شروع شماره‌گذاری از صفر است.

ورودی نمونه

```
4 2 6
I love aircrafts
Airbus A380 is one of the safest aircraft ever in the world
Boeing 777 is named as triple 7 in the UK aviation
F22 Raptor is the most adavenced air superiority fighter ever made by the united states
Where do you like to travel with a Boeing aircraft
what is the safest aircraft in the world
```

خروجی نمونه

```
2
1
```