یادگیری عمیق
نیم‌سال دوم ۹۸ــ۹۹

مدرس: دکتر سلیمانی

دانشکده‌ی مهندسی کامپیوتر

تمرین سری اول | مرور جبر خطی و مشتق‌گیری | موعد تحویل: ۲ اسفند

نمره: ۱۰ + ۳۵

مسئله‌ی ۱. مقدار و بردار ویژه (۱۰ نمره)

۱. (۱۰ نمره) ماتریس تبدیل دوران در دو بعد به اندازه‌ی θ درجه به فرم زیر است:

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

مقادیر و بردارهای ویژه‌ی این ماتریس را بدست آورید و صحت رابطه‌ی زیر را برای آن نشان دهید:

$$\det(X) = \prod_{i=1}^{n} |\lambda_i|$$

که در آن λ_iها مقادیر ویژه‌ی ماتریس X هستند.

در نهایت با استفاده از تجریه‌ی مقادیر ویژه نشان دهید:

$$R(n\theta) = R^n(\theta)$$

مسئله‌ی ۲. مشتق بردار و ماتریس (۲۵ نمره)

۱. (۲۰ نمره) موارد زیر را اثبات کنید ($y \in \mathbb{R}, x \in \mathbb{R}^n, a \in \mathbb{R}^n, X \in \mathbb{R}^{n\times n}, A \in \mathbb{R}^{n\times n}$):

الف) $\frac{\partial x^T a}{\partial x} = \frac{\partial a^T x}{\partial x} = a^T$

ب) $\frac{\partial}{\partial y}\mathrm{tr}(X) = \mathrm{tr}\left(\frac{\partial}{\partial y}X\right)$[۱]

پ) $\frac{\partial}{\partial X}\mathrm{tr}(X^T A X) = X^T(A + A^T)$

ت) $\frac{\partial}{\partial X}\log(\det(X)) = X^{-T}$

راهنمایی: معکوس یک ماتریس مربعی به شکل زیر بدست می‌آید:

$$(X^{-1})_{ij} = \frac{1}{\det{(A)}} C_{ji}$$

که در آن C ماتریس Cofactor است و داریم:

$$C_{ij} = (-1)^{i+j}\det{(M_{ij})}$$

که M_{ij} ماتریس X بدون سطر i و ستون j است.

۲. (۵ نمره) اگر x و y دو بردار به ترتیب n و m بعدی باشند و بین آن‌ها رابطه‌ی $y = f(x)$ برقرار باشد، آنگاه مشتق بردار y نسبت به بردار x را ژاکوبین گویند و به صورت ماتریس $m \times n$ زیر نمایش می‌دهند:

$$J_y(x) = \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

تابع برداری زیر را در نظر بگیرید:

$$f(x_1, x_2, x_3) = (x_1 x_2^2 {x_3}^2, {x_3}^2 \sin(x_2), {x_1}^2 e^{x_2})$$

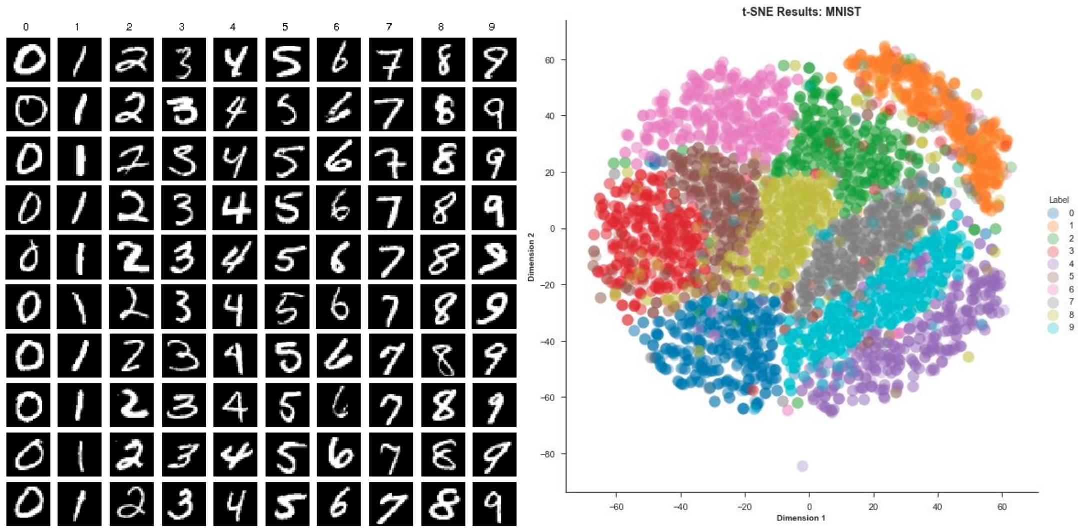
ژاکوبین تابع f را در نقطه‌ی $x = (1, \pi, -1)$ محاسبه کنید.

---

[۱] به طور کلی trace یک نگاشت خطی است و با تمام نگاشت‌های خطی دیگر همانند مشتق، انتگرال، امید ریاضی و ... قابل جابجایی است.

مسئله‌ی ۳. الگوریتم کاهش ابعاد SNE (۱۰ نمره)

۱. (۱۰ نمره) یکی از محبوب‌ترین روش‌های نمایش داده‌های با ابعاد زیاد، کاهش آن‌ها توسط الگوریتم t-SNE[2] به ۲ (یا ۳) بعد و نمایش آن بر روی مختصات دکارتی است. برای مثال شکل زیر مجموعه‌ی دادگان MNIST شامل تصاویر ۷۸۴ بعدیِ (۲۸×۲۸) رقم‌های دست‌نویس ۰ تا ۹ است که به ۲ بعد کاهش یافته‌اند:



در این قسمت نسخه‌ی اولیه‌ی این الگوریتم به نام SNE مورد استفاده قرار می‌گیرد.

این الگوریتم دارای دو فاز است. در فاز اول یک توزیع احتمالاتی بر روی هر جفت داده‌ی با ابعاد بالا تخمین می‌زند به‌گونه‌ای که احتمال انتخاب جفت‌های شبیه به هم بالا باشد. در فاز دوم نیز توزیع احتمالاتی مشابهی بر روی جفت داده‌ها با ابعاد پایین ارائه می‌دهد و سعی می‌کند برای هر جفت داده، فاصله‌ی توزیع اول با توزیع دوم یکی باشد. در این الگوریتم از دیورژانس Kullback-Leibler برای تخمین فاصله‌ی دو توزیع استفاده می‌شود. همچنین برای معیار شباهت نیز از فاصله‌ی اقلیدسیِ نقاط استفاده می‌شود که می‌توان آن را به هر معیار دیگری تغییر داد.

مجموعه‌ی N نقطه‌ی با ابعاد بالای $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N$ را در نظر بگیرید. الگوریتم SNE با استفاده از فاصله‌ی اقلیدسی بین نقطه‌ی x_i با نقطه‌ی x_j توزیع شرطی $p_{j|i}$ که در واقع احتمال انتخاب‌شدن نقطه‌ی jام به شرط انتخاب‌شدن نقطه‌ی iام است را تخمین می‌زند. در الگوریتم SNE توزیع نقاط برای انتخاب‌شدن به شرط انتخاب نقطه‌ی i، یک توزیع گوسی به مرکز x_i و واریانس σ_i^2 است که باعث می‌شود نقاط نزدیک به آن احتمال بیشتری برای انتخاب‌شدن داشته باشند. در واقع داریم:

$$p_{j|i} = \frac{e^{-\frac{\|x_j - x_i\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}}}$$

[2] t-distributed stochastic neighbor embedding

از آن‌جایی که این الگوریتم صرفاً از شباهت جفت‌داده‌ها استفاده می‌کند، مقدار $p_{i|i}$ را برابر صفر قرار می‌دهد. توجه نمایید که σ_iها هایپرپارامترهای مساله هستند که معمولاً یکی در نظر گرفته می‌شوند و ایده‌هایی برای تعیین آن‌ها وجود دارد که فراتر از هدف این تمرین است.

به طریق مشابه برای N نقطه‌ی با ابعاد کم $y_1, y_2, \ldots, y_N$ که تبدیل‌شده‌ی نقاط با ابعاد بالا هستند نیز داریم:

$$q_{j|i} = \frac{e^{\|y_j - y_i\|^2}}{\sum_{k \neq i} e^{\|y_i - y_k\|^2}}$$

به دلیل مشابه مقدار $q_{i|i}$ برابر صفر است.

می‌توان ماتریس P و Q را ماتریس شباهت بین نقاط در دو فضا تعبیر کرد که هدف الگوریتم SNE نزدیک‌کردن این دو ماتریس به یکدیگر است.

در انتها هدف الگوریتم SNE کمینه‌کردن مجموع فاصله‌ی توزیع انتخاب همه‌ی نقاط به شرط نقطه‌های دیگر است.

$$C = \sum_{i=1}^{N} KL(P_i || Q_i) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{j|i} \log \left(\frac{p_{j|i}}{q_{j|i}}\right)$$

که P_i و Q_i توزیع انتخاب داده‌های دیگر به شرط انتخاب داده‌ی iام است (با توجه به روابط $p_{j|i}$ و $q_{j|i}$ مشخص است که P_i و Q_i توزیع‌های احتمالاتی روی همه‌ی داده‌ها هستند).

توجه نمایید که فرض می‌شود $0 \times \log 0$ برابر صفر است.

در نهایت برای بدست آوردن بهترین y_iها کافی است با استفاده از روش گرادیان نزولی تابع هزینه را نسبت به آن‌ها کمینه کرد. ثابت کنید:

$$\frac{\partial C}{\partial y_i} = 2 \sum_{j=1}^{N} (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j)$$