9/11/23

Lecture 2: Tensors, Jennrich's algorithm, applications

Tensor contraction:

Given $a_1, \dots, a_\ell \in \mathbb{R}^d$, $T \in (\mathbb{R}^d)^{\otimes \ell}$,

$$T(a_1, \dots, a_\ell) = \sum_{i_1, \dots, i_\ell} T_{i_1 \cdots i_\ell} (a_1)_{i_1} \cdots (a_\ell)_{i_\ell}$$

can be interpreted as evaluating polynomial w/ coefficients given by entries of T at the point $(a_1, \dots, a_\ell)$.

Partial contraction:

$$T(a_1, \dots, a_m, :, \dots :) \in (\mathbb{R}^d)^{\otimes \ell - m}$$

$$T(a_1, \dots, a_m, :, \dots, :)_{i_1 \cdots i_{\ell-m}} = T(a_1, \dots, a_m, e_{i_1}, \dots, e_{i_{\ell-m}})$$

"higher-order" generalization of matrix-vector mult:
if T is matrix, i.e. order-2 tensor,

$$T(:, z) = Tz$$

Ex: If $T = u \otimes v \otimes w$, then

$$T(:, :, z) = (u \otimes v) \cdot \langle w, z \rangle$$

so if $T = \sum_i u_i \otimes v_i \otimes w_i$, then

$$T(:, :, z) = \sum_i (u_i \otimes v_i) \cdot \langle w_i, z \rangle$$

Jennrich's algorithm:

Suppose we are given $T = \sum_{i=1}^{k} u_i \otimes v_i \otimes w_i$.

Sample z, z' randomly from $\mathbb{S}^{d-1}$.

Form

$$M_z \triangleq T(:,:,z) = \sum_i (u_i \otimes v_i) \langle w_i, z \rangle$$

$$M_{z'} \triangleq T(:,:,z') = \sum_i (u_i \otimes v_i) \cdot \langle w_i, z' \rangle$$

Note: if $U, V, W \in \mathbb{R}^{d \times k}$ are matrices whose columns are $\{u_i\}$, $\{v_i\}$, $\{w_i\}$ respectively, and

$$D_z \triangleq \text{diag}(\langle w_1, z \rangle, \ldots, \langle w_k, z \rangle)$$

$$D_{z'} \triangleq \text{diag}(\langle w_1, z' \rangle, \ldots, \langle w_k, z' \rangle),$$

then

$$M_z = U D_z V^T$$

$$M_{z'} = U D_{z'} V^T$$

Now we use the same "simultaneous diagonalization" trick as in matrix pencil method!

$$M_2 M_{2'}^{+} = UD_2V^T(UD_{2'}V^T)^{+}$$
$$= UD_2 \underbrace{V^T (V^T)^{+}}_{=Id} D_{2'}^{-1} U^{+}$$
$= Id$, because V has full col. rank by assumption
$$= UD_2 D_{2'}^{-1} U^{+}$$

so the eigenvectors of $M_2 M_{2'}^{+}$ are the columns of U! i.e. we get $\{u_i\}$

Similarly,

$$M_2^{+} M_{2'} = (UD_2V^T)^{+} UD_{2'}V^T$$
$$= (V^T)^{+} D_2^{-1} \underbrace{U^{+}U}_{=Id} D_{2'} V^T$$
$= Id$, because U has full col. rank by assumpt.
$$= (V^T)^{+} D_2^{-1} D_{2'} V^T$$

so eigenvectors of $(M_2^{+} M_{2'})^T$ are the columns of V, i.e we get $\{v_i\}$

Q: How to "pair up" $\{u_i\}$ and $\{v_i\}$? (we only know them up to permutation.)

A: Eigenvalues of $M_2 M_{2'}^+$ and $(M_2^+ M_{2'})^T$ are in 1-1 correspondence:

those of $M_2 M_{2'}^+$ are $\left\{ \frac{\langle w_i, z\rangle}{\langle w_i, z'\rangle} \right\}_{i=1,\dots,k}$

reciprocals

those of $(M_2^+ M_{2'})^T$ are $\left\{ \frac{\langle w_i, z'\rangle}{\langle w_i, z\rangle} \right\}_{i=1,\dots,k}$

N.B.: Here, we use randomness of z, z' to ensure that $\langle w_i, z\rangle, \langle w_i, z'\rangle \neq 0$, and together w/ non-collinearity of w_i's, that there are no "accidental" reciprocals so eigenvalues really are in 1-1 correspondence as claimed.

Thus, can recover $\{u_i \otimes v_i\}_{i=1,\dots,k}$.

Finally, need to recover $w_1, \dots, w_k$.

Idea. set up a linear system

Note, $\underbrace{T_{abc}}_{\text{known}} = \sum_i \underbrace{(u_i)_a}_{\text{known}} \underbrace{(v_i)_b}_{\text{known}} \underbrace{(w_i)_c}_{\text{unknown}}$ $(*)$

This gives d^3 equations in kd unknowns.
Need to show the equations have a <u>unique</u> soln.

Let $\lambda^{(ab)} \in \mathbb{R}^k$ have entries $\{(u_i)_a (v_i)_b\}_{i \in 1,\dots,k}$.

Then $(*) \Leftrightarrow T_{abc} = \langle \lambda^{(ab)}, W_{c:} \rangle$,

where $W_{c:}$ denotes c-th row of W.

To show uniqueness of soln., the following suffices:

<u>Lem</u>: $\{\lambda^{(ab)}\}_{a,b \in [d]}$ spans $\mathbb{R}^k$

<u>Pf</u>: Let $\Lambda \in \mathbb{R}^{d^2 \times k}$ have columns $\{\lambda^{(ab)}\}$.

Because $d^2 > k$, suffices to show Λ has full column rank.

Suppose to contrary that $\exists\, c \in \mathbb{R}^k$ s.t.

$$\sum_i c_i \Lambda_i = \vec{0} \qquad (**)$$

Note $\Lambda_i = \underbrace{\mathrm{vec}(u_i \otimes v_i)}_{\substack{\text{i.e. flatten matrix} \\ u_i \otimes v_i \text{ into vector}}}$

So (★★) implies

$$\sum_i c_i u_i v_i^T = 0$$

Suppose WLOG $c_1 \neq 0$.

Let x be s.t. $\langle u_1, x \rangle \neq 0$, yet $\langle u_i, x \rangle = 0 \ \forall i > 1$.
(exists b/c $u_1, \ldots, u_k$ linearly independent)

Then

$$\underbrace{x^T \sum_i c_i u_i v_i^T}_{\substack{\| \\ c_1 \langle u_1, x \rangle v_1^T \neq 0}} = 0$$

, contradiction. $\square$

Q1: What about higher-order tensors?

e.g. $T = \sum_{i=1}^{k} u_i \otimes v_i \otimes w_i \otimes x_i$

A "reshape" T into 3^{rd}-order tensor, e.g. take

$$T' = \sum_{i=1}^{k} \underbrace{\mathrm{vec}(u_i \otimes v_i)}_{\mathbb{R}^{d^2}} \otimes w_i \otimes x_i$$

To apply Jennrich's to T', need

1). $\{u_i \otimes v_i\}$ are linearly indep

2). $\{w_i\}$ linearly indep

3). x_i, x_j non-collinear for $i \neq j$

note: by lemma above, 1) holds if $\{u_i\}$ linearly indep and $\{v_i\}$ linearly indep

---

Q2: What if

$$T = \sum u_i \otimes v_i \otimes w_i + \boxed{\text{noise}} \;?$$

A: instead of U, V being full column rank, need that $\sigma_{min}(U)$, $\sigma_{min}(V)$ not too small

See Pset 1 Q#2

APPLICATIONS:

① MIXTURES OF EXPONENTIALS:

Recall from lecture 1:

get access to

$$G: \omega \mapsto \frac{1}{k} \sum_{j=1}^{k} e^{2\pi i \langle \omega, \mu_j \rangle}$$

for any $\|\omega\| \leq 1$ where $\omega, \mu_1, \dots, \mu_k \in \mathbb{R}^2$

Goal: recover $\mu_1, \dots, \mu_k$.

for m sufficiently large

Alg: 1) Sample $w_1, \dots, w_{m} \in \underbrace{B(0.49)}_{\text{2D ball of radius 0.49 around origin}}$

$v, v' \sim \mathbb{S}^1$

2) Define $x_1 = 0.02 \cdot v$, $x_2 = 0.02 \cdot v'$

3) Define $T \in \mathbb{R}^{m \times m \times 2}$

by $T_{abc} = G[\underbrace{w_a + w_b + x_c}_{\|\cdot\| \leq 1 \text{ by design}}]$

Then $$T_{abc} = \frac{1}{k}\sum_{j=1}^{k} \underbrace{e^{2\pi i\langle w_a, \mu_j\rangle}}_{(u_i)_a} \cdot \underbrace{e^{2\pi i\langle w_b, \mu_j\rangle}}_{(u_i)_b} \cdot \underbrace{e^{2\pi i\langle x_c, \mu_j\rangle}}_{(w_i)_c}$$

where $u_i \in \mathbb{R}^d$ and $w_i \in \mathbb{R}^2$

So $T = \frac{1}{k}\sum_{j=1}^{k} u_i \otimes u_i \otimes w_i$, and we can apply Jennrich's provided $\{u_i\}$ linearly independent and w_i's non-collinear.

<u>Note</u>: This is essentially a generalization of matrix pencil method to higher dimensions!

---

② <u>SPHERICAL GAUSSIAN MIXTURES</u>:

<u>Unknown</u>: $\mu_1, \dots, \mu_k \in \mathbb{R}^d$ linearly indep.

$\lambda_1, \dots, \lambda_k \in [0,1]$ s.t. $\sum_j \lambda_j = 1$

<u>Given</u>: iid samples from mixture model q, where

$$q = \sum_{i=1}^{k} \lambda_i \cdot N(\mu_i, Id)$$

i.e. to sample from q:

1). sample $i \in [k]$ w.p. λ_i

2). Sample $g \sim N(0, I_d)$

3). output $\mu_i + g$

<u>Goal</u>: estimate $\{\mu_i\}$, $\{\lambda_i\}$ up to small error

<u>Goal</u>: estimate $\mu_1, \ldots, \mu_k$ up to error

<u>Alg</u>: $\mathbb{E}[x] = \sum_{i=1}^{k} \lambda_i \mathbb{E}[\mu_i + g] = \sum_{i=1}^{k} \lambda_i \mu_i$

$$\mathbb{E}[x^{\otimes 3}] = \sum_{i=1}^{k} \lambda_i \mathbb{E}[(\mu_i + g)^{\otimes 3}]$$

$\mathbb{E}[g^{\otimes 3}] = 0$ by symmetry

$\mathbb{E}[\text{" "}] = 0$ by symmetry

$$= \sum_{i=1}^{k} \lambda_i \mathbb{E}\Big[\mu_i^{\otimes 3} + g^{\otimes 3} + \boxed{\mu_i \otimes \mu_i \otimes g + \mu_i \otimes g \otimes \mu_i + g \otimes \mu_i \otimes \mu_i} + \underbrace{\mu_i \otimes g \otimes g}_{} + g \otimes g \otimes \mu_i + g \otimes \mu_i \otimes g\Big]$$

$\mathbb{E}[\text{" "}] = I_d$, and similarly

$$= \sum_{i=1}^{k} \lambda_i \mu_i^{\otimes 3} + \Big(\sum_{i=1}^{k} \lambda_i \mu_i\Big) \otimes_3 I_d,$$

where $z \otimes_3 I_d \triangleq \sum_{a=1}^{d} z \otimes e_a \otimes e_a + e_a \otimes z \otimes e_a + e_a \otimes e_a \otimes z$

So $\mathbb{E}[x^{\otimes 3}] - \mathbb{E}[x] \otimes_3 Id = \sum_{i=1}^{k} \lambda_i \mu_i^{\otimes 3}$

Can run Jennrich's to learn the parameters.

---

③ INDEPENDENT COMPONENT ANALYSIS (BONUS)

Unknown: $A \in \mathbb{R}^{d \times d}$

Given: iid samples of the form

$$z = Ax + g \quad , \quad x \sim q \; , \; g \sim N(0, Id)$$

for a product distribution q

interpretation: q generates d independent signals, e.g. conversations at a dinner party, and we and we only observe noisy combinations of these signals, e.g. picked up by mics in the room. Want to unscramble by learning A.

Idea: $\mathbb{E}[z_i z_j z_k z_\ell]$ unwieldy to write down (try it yourself), so use cumulants instead

(see slides for background on cumulants)

$$K(z_i, z_j, z_k, z_l)$$

$$= K\Big(\sum_s A_{is} x_s + g_i,\ \sum_s A_{js} x_s + g_j,$$

$$\sum_s A_{ks} x_s + g_k,\ \sum_s A_{ls} x_s + g_l\Big)$$

(Gaussian cumulants vanish)

$$= K\Big(\sum_s A_{is} x_s, \ldots, \sum_s A_{ls} x_s\Big)$$

(additivity)

$$= \sum_{s_1, \ldots, s_4} K\big(A_{is_1} x_{s_1}, \ldots, A_{ls_4} x_{s_4}\big)$$

(independent cumulants vanish)

$$= \sum_s K\big(A_{is} x_s, \ldots, A_{ls} x_s\big)$$

$$= \sum_s A_{is} A_{js} A_{ks} A_{ls} \cdot \underbrace{K(x_s, x_s, x_s, x_s)}_{\triangleq \lambda_s}$$

so if $T_{ijkl} = K(z_i, z_j, z_k, z_l)$, then

$$T = \sum_s \lambda_s A_{:s}^{\otimes 4}$$

note $\lambda_s = \underset{x}{\mathbb{E}}[x_s^4 - 3]$

If $g = N(0, Id)$, then $\lambda_s = 0$, but this case is actually impossible!

$$z = Ax + g \sim N(0, Id + A^TA),$$

so would only be able to recover A up to rotation of its rows...

But if g "non-Gaussian" so $\lambda_s \neq 0$, then Jennrich's can be used to recover A provided A has full rank!