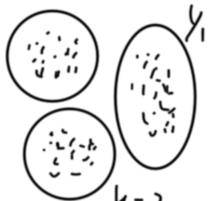
Lecture 9

Clustering / GMMs via Tensor decomposition



$k=3$

$y_1, \cdots, y_n \in \mathbb{R}^d$

observe. $\underline{y}_r = \mu_{z^*(r)}^* + \underline{w}_r$, $\underline{w}_r \sim N(0, I_d)$ * independent

unknown: $\mu_1^*, \cdots, \mu_k^* \in \mathbb{R}^d$, all linearly independent

$z^*(r) \in \{1, \cdots, k\}$

goal: recover μ_i^*

$$\min_{\substack{z(r) \in \{1,\cdots,k\} \\ \mu \in \mathbb{R}^d}} \sum_{i=1}^n \| Y_r - \mu_{z(r)} \|_2^2$$

- Initialize $\mu_1, \cdots, \mu_n$
- Iterate
  - for each point, set $z(r)$, such that Y_r is assigned to closest M.
  - Update $\mu_i = \frac{1}{S_i} \sum_{\substack{r=1 \\ z(r)=i}} Y_r$, $S_i = |\{ r \mid z(r) = i \}|$

Moments

$\underline{X} \in \mathbb{R}$, $\mathbb{E}[\underline{X}^l]$ $\forall l \in \mathbb{N}$

$\underline{X} \in \mathbb{R}^d$, $\mathbb{E}[\underline{X}] \in \mathbb{R}^d$

$\mathbb{E}[\underline{X}\underline{X}^T]$ contains $\mathbb{E}[X_i X_j]$ for all i, j $\quad$ $\underline{X}\underline{X}^T \triangleq \underline{X}^{\otimes 2}$

$(\underline{X}^{\otimes 3})_{i,j,k \in [1,d]}$, $(\underline{X}^{\otimes 3})_{ijk} = \underline{X}_i \cdot \underline{X}_j \cdot \underline{X}_k$

"lives in d^3 dimensions"

$\underline{X}^{\otimes l}$, $(i_1, i_2, \cdots, i_l) \to \underline{X}_{i_1} \cdot \underline{X}_{i_2} \cdot \cdots \cdot \underline{X}_{i_l}$

lives in d^l dimensions

The l-th moment tensor: $\mathbb{E}[\underline{X}^{\otimes l}]$

Distribution D, input $\underline{X}_1, \cdots, \underline{X}_n \overset{iid}{\sim} D$

$\hat{M}_l = \frac{1}{n} \sum_{i=1}^n \underline{X}^{\otimes l}$

Tensor rules

- $(a+b)^{\otimes 2} = a^{\otimes 2} + b^{\otimes 2} + a \otimes b + b \otimes a$
- $(a \otimes b)_{ij} = a_i \cdot b_j$
- $(a_1 \otimes \cdots \otimes a_l)_{i_1, i_l} = (a_1)_{i_1} \cdot \cdots \cdot (a_l)_{i_l}$
- $(a+b)^{\otimes 3} = (a+b) \otimes (a+b)^{\otimes 2}$

$$= (a+b) \otimes (a^{\otimes 2} + b^{\otimes 2} + a \otimes b + b \otimes a)$$
$$= a^{\otimes 3} + a \otimes b^{\otimes 2} + a^{\otimes 2} \otimes b + a \otimes b \otimes a + b \otimes a^{\otimes 2} + b^{\otimes 3} + b \otimes a \otimes b + b^{\otimes 2} \otimes a$$

Input: $Y_1, \cdots, Y_n$

$$M_1 = \mathbb{E}_{Y_1, \cdots, Y_n}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right]$$
$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n M_{z^*(i)} + W_i\right]$$
$$= \frac{1}{n} \sum_{i=1}^n \mu_i(r) + \underbrace{\frac{1}{n} \sum_{i=1}^n W_i}_{=0}$$
$$= \sum_{j=1}^k \alpha_j \mu_j$$

* $\alpha_j = \frac{|\{ r \mid z^*(r) = j \}|}{n}$

$$M_2 = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i^{\otimes 2}\right]$$
$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (\mu_{z(i)} + \underline{W}_i)^{\otimes 2}\right]$$
$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \mu_{z(i)}^{\otimes 2} + \underline{W}_i^{\otimes 2} + \mu_{z(i)} \otimes W_i + W_i \otimes \mu_{z_i^*}\right]$$
$$= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[\mu_{z(i)}^{\otimes 2} + \cancel{\underline{W}_i^{\otimes 2}} + \cancel{\mu_{z(i)} \otimes W_i} + \cancel{W_i \otimes \mu_{z_i^*}}\right]$$

0 $\quad$ 0 $\quad$ 0

$\mu_{z^*(i)}^{\otimes 2}$ $\quad$ $[\mathbb{E}[W_i^{\otimes 2}]]_{kc}$ $\quad$ 0

Tensor rules (continued)

$$a\otimes b + a\otimes c = a\otimes(b\otimes c)$$
$$\sum_{i=1}^{n} a\otimes b_i = a\otimes\left(\sum_{i=1}^{n} b_i\right)$$

$$= \mathbb{E}[W_i]_r\cdot\mathbb{E}[W_i]_s$$
$$= \begin{cases}\mathbb{E}[W_i]_r^2 = 1, & r=s\\ \mathbb{E}[W_i]_r\mathbb{E}[W_i]_s = 0, & r\neq s\end{cases}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mu_{z^*(i)}^{\otimes 2} + \sum_{r=1}^{d} e_r^{\otimes 2}$$
$$= \sum_{j=1}^{k}\alpha_j\mu_j^{\otimes 2} + \sum_{r=1}^{d} e_r^{\otimes 2}$$

$$M_3 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\underline{Y_i}}\left[\left(\mu_{z^*(i)}^* + \underline{W_i}\right)^{\otimes 3}\right]$$

$a = \mu_{z^*(i)}^*$, $b = \underline{W_i}$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mu_{z^*(i)}^{*\otimes 3}}_{=\sum_{j=1}^{k}\alpha_j\mu_j^{*\otimes 3}} + \underbrace{\mathbb{E}\Big[\overset{(A)}{\mu_{z^*(i)}^*\otimes W_i^{\otimes 2}} + \overset{(B)}{W_i\otimes\mu_{z^*(i)}^*\otimes W_i} + \overset{(C)}{W_i^{\otimes 2}\otimes\mu_{z^*(i)}^*}\Big]} + \overset{0}{\cancel{\mathbb{E}[\underline{W_i}^{\otimes 3}]}}$$

lives in d^3 dimensions

$$\left(\mathbb{E}[W_i^{\otimes 3}]\right)_{r,s,t} = \mathbb{E}\left[(W_i)_r(W_i)_s(W_i)_t\right] = \begin{cases}0, & r,s,t \text{ are all disjoint}\\ 0, & \text{if two are equal}\\ 0, & r=s=t\end{cases}$$

$\mathbb{E}[W_i] = -\mathbb{E}[W_i]$

$$(A) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mu_{z^*(i)}^*\otimes W^{\otimes 2}\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[\mu_{z^*(i)}^*\otimes\left(\sum_{r=1}^{d}e_r^{\otimes 2}\right)\right]$$
$$= \left[\frac{1}{n}\sum_{i=1}^{n}\mu_{z^*(i)}^*\right]\otimes\left(\sum_{r=1}^{d}e_r^{\otimes 2}\right) = \left(\sum_{j=1}^{k}\alpha_j\mu_j^*\right)\otimes\left(\sum_{r=1}^{d}e_r^{\otimes 2}\right)$$

$$(C) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W^{\otimes 2}\right]\otimes\mu_{z^*(i)}^*$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{r=1}^{d}e_r^{\otimes 2}\right)\otimes\mu_{z^*(i)}^*$$
$$= \left(\sum_{r=1}^{d}e_r^{\otimes 2}\right)\otimes\underbrace{\sum_{j=1}^{k}\alpha_j\mu_j^*}_{=T_1}$$

$$(B) = \sum_{r=1}^{d}e_r\otimes T_1\otimes e_r$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_i\otimes\mu_{z^*(i)}^*\otimes W_i\right]$$
$$= \sum_{j=1}^{k}\alpha_j\underbrace{\mathbb{E}\left[\underline{W}\otimes\mu_j^*\otimes\underline{W}\right]}_{=\sum_{r=1}^{d}e_r\otimes\mu_j^*\otimes e_r}$$
$$= \sum_{j=1}^{k}\alpha_j\left(\sum_{r=1}^{d}e_r\otimes\mu_j^*\otimes e_r\right)$$
$$= \sum_{r=1}^{d}\sum_{j=1}^{k}\alpha_j\, e_r\otimes\mu_j^*\otimes e_r$$
$$= \sum_{r=1}^{d}e_r\otimes T_1\otimes e_r$$

$$M_3 = \sum_{j=1}^{k}\alpha_j\mu^{\otimes 3} + \underbrace{T_1\otimes\left(\sum_{r=1}^{d}e_r^{\otimes 2}\right) + \sum_{r=1}^{d}e_r\otimes T_1\otimes e_r + \left(\sum_{r=1}^{d}e_r^{\otimes 2}\right)\otimes T_1}_{\text{unknown}}$$

* Given M_1, M_3, we can compute $\sum_{j=1}^{k}\alpha_j\cdot\mu^{\otimes 3}$.

$(a_1\otimes a_2\otimes I_d)\cdot(b_1\otimes b_2\otimes b_3) = \langle a_1,b_1\rangle\cdot\langle a_2,b_2\rangle\cdot b_3$,

$a^{\otimes 2}\cdot b^{\otimes 2} \Rightarrow (a_1\otimes\cdots\otimes a_\ell)\cdot(b_1\otimes\cdots\otimes b_\ell) = \langle a_1,b_1\rangle\cdots\langle a_\ell,b_\ell\rangle$

* $a^T(bb^T)a = \langle a,b\rangle^2$

Claim 1: "sufficient to recover up to scaling"

Proof: Assume μ_j^* are orthogonal

For $i=1\cdots k$, $\hat{\mu}_i = \underset{\in\mathbb{R}}{c}\cdot\mu_i^*$, $\quad \hat{\mu}_i^T\left(\sum_{j=1}^{n}\alpha_j\cdot\mu_j^*\mu_j^{*T}\right)\mu_i = \sum_{j=1}^{k}\alpha_j\langle\hat{\mu}_j,\mu_j^*\rangle^2 = \alpha_1\cdot\langle c_i\mu_i^*,\mu_i^*\rangle^2$

$$= c_i^2\alpha_i\cdot\|\mu_i^*\|^4 = c_i^2\alpha_1\cdot\|\mu_i^*\|^4 = \frac{\alpha_4\|\mu_1\|^4}{c_i^2}$$

$\|\mu_i^*\| = \frac{1}{c_i}\cdot\|\hat{\mu}_i\|$

* $(\hat{\mu}_i^{\otimes 2})\cdot\left(\sum_j\alpha_j\mu_j^{\otimes 2}\right)$

$$= \sum_j \alpha_j \left[(\hat{\mu}_1^{\otimes 3}) - (\mu_j^{*\otimes 3}) \right]$$

$$= \sum_j \alpha_j \langle \hat{\mu}_1, \mu_j^* \rangle^3 = \alpha_1 \cdot \langle c_1 \cdot \mu_1^*, \mu_1^* \rangle^3 = \alpha_1 \cdot c_1^3 \cdot \|\mu_1^*\|^6$$

$$\rightarrow \beta_1 = \frac{\alpha_1}{c_1^2}, \ \beta_2 = \frac{\alpha_1}{c_1^3}$$

$$\rightarrow \frac{\beta_1}{\beta_2} = c_1$$

Eigen decomposition

$M \in \mathbb{R}^{d \times d}$, Λ diag, $P \succ t$, $M = P \Lambda P^{-1}$

columns of P have unit norm

elements of Λ are all distinct

$\rightarrow$ decomposition is unique up to permutation & sign of P

Theorem

$a_1 \cdots a_d$	lin. indep.	Given $T = \sum_{i=1}^{d} a_i \otimes b_i \otimes c_i$, can recover $\{a_i\}, \{b_i\}, \{c_i\}$ up to
$b_1 \cdots b_d$	lin. indep.	permutational (signed) scaling.
$c_1 \cdots c_d$	lin. indep.	$a_i = \alpha_i \cdot \mu_i^* \cdot b_i = \mu_i^*, \ c_i = \mu_i^*$

$g_1 \sim N(0, I_d)$, $g_2 \sim N(0, I_d)$

$$\underline{M}_1 = (I_d \otimes I_d \otimes g_1) = \sum_{i=1}^{r} \langle g_1, c_i \rangle \cdot a_i \otimes b_i = \sum_{i=1}^{r} \langle g_1, c_i \rangle a_i b_i^T$$

$$= \underset{\text{cols } a_i}{A} \ \underset{\text{diag } \langle g_1, c_i \rangle}{\underline{\Lambda}_1} \ \underset{\text{cols } b_i}{B}$$

$$\underline{M}_2 = (I_d \otimes I_d \otimes g_2) = \sum_{i=1}^{r} \langle g_2, c_i \rangle a_i b_i^T = A \underset{\text{diag } \langle g_2, c_i \rangle}{\underline{\Lambda}_2} B$$

$$\underline{M}_1 \underline{M}_2^{-1} = (A \Lambda_1 B)(A \Lambda_2 B)^{-1}$$

$$= A \Lambda_1 \cancel{B B^{-1}} \Lambda_2^{-1} A^{-1}$$

$$= A \underbrace{\Lambda_1 \Lambda_2^{-1}}_{\text{diag } \frac{\langle g_1, c_i \rangle}{\langle g_2, c_i \rangle}} A^{-1}$$

all distinct.

$\longrightarrow$ eigen decomp. of $\underline{M}_1 \underline{M}_2$ has (scaled) a_i as eigenvectors

$\longrightarrow$ recover up to scale

$\underline{M}_2^T (\underline{M}_1^{-1})^T \longrightarrow$ can recover b_i up to scaling

$(c_1)_1 \cdots (c_r)_1$

$$\hookrightarrow \underset{\text{vector in } \mathbb{R}^{d^2}}{L} = \sum_{i=1}^{r} \underbrace{(a_i \otimes b_i)}_{\text{vectors in } \mathbb{R}^{d^2}} \cdot (c_i)_1$$

*can solve for (c_1) if $(a_i \otimes b_i)$ are lin. indep.

$$\lambda_1 \cdot \begin{pmatrix} a_{11} \cdot b_1 \\ \vdots \\ a_{12} \cdot b_1 \\ \vdots \\ a_{1d} \cdot b_1 \\ \vdots \end{pmatrix} + \lambda_2 \begin{pmatrix} a_{21} \cdot b_2 \\ \vdots \\ a_{22} \cdot b_2 \\ \vdots \\ a_{2d} \cdot b_2 \\ \vdots \end{pmatrix} + \cdots + \lambda_r \begin{pmatrix} a_{r1} \cdot b_r \\ \vdots \\ a_{r2} \cdot b_r \\ \vdots \\ a_{rd} \cdot b_r \\ \vdots \end{pmatrix}$$