03.03.2020

Result: $B = A^T A$ $\quad A = \sum \sigma_i u_i v_i^T \Rightarrow \sum_i \sigma_i^2 v_i v_i^T = B$

$$B^k = \sum_i \sigma_i^{2k} v_i v_i^T \approx \sigma_1^{2k} v_1 v_1^T$$

Sparse Matrix ~ # of non-zero elts

If A is a sparse matrix, $A^T A$ need not to be sparse matrix.

Theorem. $A_{n \times d}$, X is a unit vector in R^d: $\underline{\langle x, v_1 \rangle} \geq \delta > 0$

$V = \text{span}\{v_1, \ldots, v_n\}$. $\sigma_m \geq (1-\varepsilon)\sigma_1$ for some $\varepsilon > 0$.

Let W be the unit vector after k-iterations $\left(k \geq \ldots \right)$

$\Downarrow$

$$k \geq \frac{\ln\left(\frac{1}{\varepsilon \delta}\right)}{2\varepsilon}$$

Power Method:

$$W = \frac{(A^T A)^k x}{\|(A^T A)^k x\|}$$ Then $\underbrace{\|W \perp V\|}_{} \leq \varepsilon$

most of the comps of W lie on that space where singular values are.

Proof: $\underset{n \times d}{A} = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

$\|x\| = 1$. $\sum_{i=1}^{d} c_i v_i$

$$(A^T A)^k x = \sum_i \sigma_i^{2k} c_i v_i \qquad |c_1| = \langle x, v_1 \rangle \geq \delta$$

$\sigma_1, \sigma_2, \ldots \sigma_m \geq (1-\varepsilon)\sigma_1$

$\sigma_{m+1}, \ldots \sigma_d \geq (1-\varepsilon)\sigma_1$

$$\Rightarrow |(A^T A)^k x|^2 = \sum_{i=1}^{d} \sigma_i^{4k} c_i^2 \geq$$

$$\rightarrow \geq \sigma_1^{4k} c_1^2 \geq \sigma_1^{4k} \delta^2$$

$$\left|(A^TA)^k x \perp V\right|^2 = \sum_{i=m+1}^{d} \sigma_i^{4k} c_i^2 \leq (1-\varepsilon)^{4k} \sigma_1^{4k} \sum_{i=m+1}^{d} c_i^2$$

$$\leq (1-\varepsilon)^{4k} \sigma_1^{4k}$$

$$\frac{(1-\varepsilon)^{2k} \cdot \sigma_1^{2k}}{\sigma_1^{2k} \delta} = \frac{(1-\varepsilon)^{2k}}{\delta}$$

$$\frac{(1-\varepsilon)^{2k} \sigma_1^{2k}}{\sigma_1^{2k} \delta} = \frac{(1-\varepsilon)^{2k}}{\delta} \leq \frac{e^{-2k\varepsilon}}{\delta} = \varepsilon$$

$$W = \frac{(A^TA)^k x}{\|(A^TA)^k x\|} \leq \frac{(A^TA)^k x}{\sigma_1^{2k} \delta}$$

$$\|W \perp V\|^2 \leq \frac{(1-\varepsilon)^{4k} \sigma_1^{4k}}{\sigma_1^{4k} \delta^2}$$

Clustering. m points $\{A_{(1)}, \dots A_{(m)}\}$ in a d-dimensional space.

Need to find "k" $= P \equiv \{P_{(1)}, \dots P_{(k)}\} \in R^d$:

$$\underline{\underline{\min}}\ d_A(P) = \min \sum_{i=1}^{m} d(A_{(i)}, P)^2$$

$$P_{(j)} : j = 1, \dots, k \qquad S_j = j = 1 \dots k \qquad \bigcup_j S_j = A$$

$$\underset{\text{j-th cluster}}{S_j} = \{A_{(j_1)}, \quad A_{(j_r)}\}$$

$$\sum_{i=1}^{r} \| A_{(j_i)} - B \|^2 = \sum_{i=1}^{r} \| A_{(j_i)} - \bar{A}_j \|^2 + r \| B - \bar{A}_j \|^2$$

Defn: Let be given a prob. distr. $P(\cdot)$ in $\mathbb{R}^d$. then the best fit line for $P(\cdot)$ in the direction v: $v = \underset{\|v\|=1}{\operatorname{argmax}} \underset{X \sim P}{E}[\langle v, x \rangle^2]$

Result: The best-fit 1-D subspace (line) for a Gaussian with (μ, σ) in $\mathbb{R}^d$ is given by $v = \mu$.

Proof: Choose $X \sim P(\cdot)$. Let $v : \|v\| = 1$

$$E_{X \sim P(\cdot)}\left[\langle x, v \rangle^2\right] = E\left[\left(\langle (x-\mu), v \rangle + \langle \mu, v \rangle\right)^2\right]$$

$$= E\left[\langle (x-\mu), v \rangle^2\right] + E\left[\langle \mu, v \rangle^2\right] = \sigma^2 + \langle \mu \cdot v \rangle^2$$

Defn: ... then k-dim best-fit subspace V_k is

$$V_k \equiv \underset{V}{\operatorname{argmax}} \; E_{X \sim P(\cdot)}\left[|X \perp V|^2 \right] \qquad \dim(V) = k.$$

Result: Any k-dim subspace $V_k \ni \mu$ is a best-fit subspace for Gaussians.

Proof: Suppose $\mu = 0$ then ok. Suppose $\mu \neq 0$, the best fit line $v = \mu$. Proceed as in SVD

Result: Suppose $P(\cdot) \sim d$-dim Gaussian (μ, σ^2) then $P(\cdot) \perp V_k$ is also Gaussian with σ^2.

k is optimal SVD subspace.

Follow the approach of SVD.

* 1. Rotate the coordinates so that $V = \text{span}\{e_1, \dots, e_k\}$ (spherically symmetric)

2. The gaussian remains spherical (σ^2) but coordinates of μ changes ($\mu = (\mu', \mu'')$)

3. $x = (x_1, \dots, x_d)$: $x' \equiv (x_1, x_2, \dots x_k)$ $x'' \equiv (x_{k+1}, \dots x_d)$

4. $[\mathcal{P}(\cdot) \perp V]$ at $(x_1, \dots x_k)$ is $e^{-\frac{\|x'-\mu'\|^2}{2\sigma^2}} \underbrace{\int_{x''} e^{-\frac{\|x''-\mu''\|^2}{2\sigma^2}}}_{=1}$

$$\frac{1}{\sigma^{d-k}(\sqrt{2\pi})^{d-k}} e^{-\frac{\|x-\mu\|^2}{2\sigma^2}}$$

$\mu = (\mu', \mu'')$

$x = (x', x'')$ – orthogonal to each other since they lie in orthogonal spaces.

Theorem: The k-dim SVD subspace V_m for a mixture of m Gaussian with $\text{span}\{\mu_1, \dots, \mu_n\} \subset V_m$

$\Rightarrow$ For a mixture of m Gaussians V_m contains the centers.

In particular, if $(c_1\mu_1 + c_2\mu_2 + \dots + c_u\mu_u = 0 \Rightarrow c_1 = c_2 = \dots = c_m = 0)$

then, $\text{span}\{\mu_1, \mu_2, \dots \mu_n\}$ is the best fit.

$$p(\cdot) \sim \sum_{i=1}^{m} w_i P_i(\cdot) \qquad \sum w_i = 1, \ w_i > 0.$$