Mixture Models and SVD

Monday, August 30, 2021 6:57 AM

Gaussian Mixture Model: $N(\mu_1, \Sigma_1)$, $N(\mu_2, \Sigma_2)$ ....

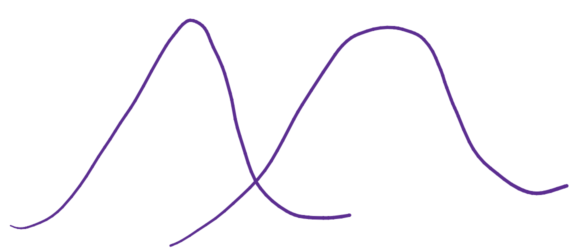
$\omega_1 \geq 0$ $\quad \omega_2 \geq 0$ $\quad \dots \omega_k$

$$\sum_{i=1}^{k} w_i = 1.$$

<u>Problem</u>: Given random samples from an unknown k-GMM estimate its parameters.

$k=1$: $\quad w_1 = 1$, μ = Sample mean $\quad \Sigma$ = sample covariance

$k=2$?



<u>special case</u>: Separable GMMs.

The components are pairwise separated.

How to measure separation.

<u>1-dim</u> $\quad |\mu_i - \mu_j| > ?? \max\{\sigma_i, \sigma_j\}$

<u>d-dim</u>. (Geometric)

$d(F_i, F_j) \geq 1 - \varepsilon.$

(probabilistic) $d_{TV}(F_i, F_j) \geq 1-\varepsilon$.

$$\boxed{d_{TV}(p,q) = \frac{1}{2}\int |p(x)-q(x)|\,dx}$$

Lemma. In 1-dim, d_{TV} "large"

$\Rightarrow$ either $\|\mu_i - \mu_j\|$ is large

or $\max\left\{\frac{\sigma_i}{\sigma_j}, \frac{\sigma_j}{\sigma_i}\right\}$ is large.

Thm. (Concentration)

$X \sim N(\mu, \sigma^2 I)$

$$\Pr\left(|X-\mu| > t\sigma\sqrt{d}\right) \leq e^{-\frac{dt^2}{2}}$$

$$\Pr\left(\left||X-\mu|^2 - d\sigma^2\right| > t\sigma^2\sqrt{d}\right) \leq 2e^{-t^2/8}$$

How to solve P1?

In the separable case, cluster and estimate each component separately.

P2. Cluster a sample from a K-GMM by component of origin.

try component of origin.

<u>How?</u> Suppose mean separated

$X, Y \in F_i$ $\quad \mathbb{E}(\|X-Y\|^2) = \mathbb{E}(\|X-\mu_i - (Y-\mu_i)\|^2)$

$$= \mathbb{E}(\|X-\mu_i\|^2) + \mathbb{E}(\|Y-\mu_i\|^2) + 0.$$

$$= 2d\sigma^2.$$

$X \in F_i, Y \in F_j$ $\quad \mathbb{E}(\|X-Y\|^2) = \mathbb{E}(\|X-\mu_i - (Y-\mu_j) + \mu_i - \mu_j\|^2)$

$$= \mathbb{E}(\|X-\mu_i\|^2) + \mathbb{E}(\|Y-\mu_j\|^2) + \|\mu_i-\mu_j\|^2$$

$$= 2d\sigma^2 + \|\mu_i-\mu_j\|^2.$$

By conc. with prob $\geq 1-2e^{-t^2/8}$

$\|X-Y\|^2 \leq 2d\sigma^2 + 2t\sqrt{d}\,\sigma^2$ $\quad X, Y \in F_i$

$\|X-Y\|^2 > 2d\sigma^2 + \|\mu_i-\mu_j\|^2 - 2t\sqrt{d}\,\sigma^2.$ $\quad X \in F_i, Y \in F_j$

$\therefore$ it suffices to have

$$\|\mu_i-\mu_j\|^2 > 4t\sqrt{d}\,\sigma^2$$

Gaussian are

$\|\mu_i - \mu_j\|^2 > 4t\sqrt{d}\,\sigma$

to ensure that pairs from same Gaussian are closer.

Cluster using distances

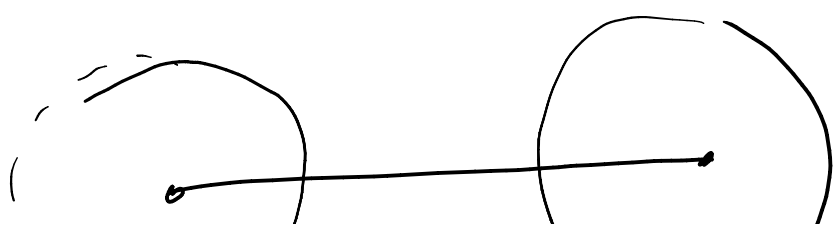
- put nearest pair in same cluster
- Repeat till only k clusters.

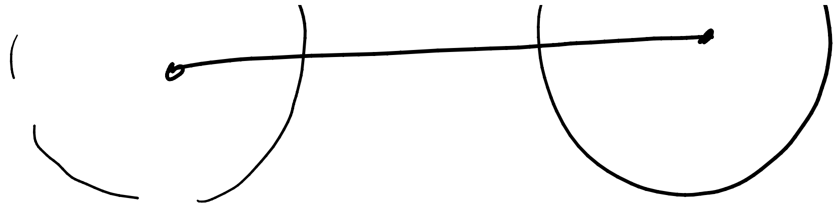
<u>Thm</u>. With prob $1-\delta$, random sample with m points and $\|\mu_i - \mu_j\| > C\left(\log\frac{m}{\delta}\cdot d\right)^{\frac{1}{4}} \max\{\sigma_i, \sigma_j\}$ can be clustered using pairwise distances in polynomial time.

<u>Pf</u>. Set $t = \sqrt{c \log \frac{m}{\delta}}$.

Is this the right answer?
Separation grows with $d^{1/4}$.

No!





Project to line joining μ_i, μ_j.

Separation needed is $O(\sigma)$. Not $\underline{d^{1/4}\sigma}$.

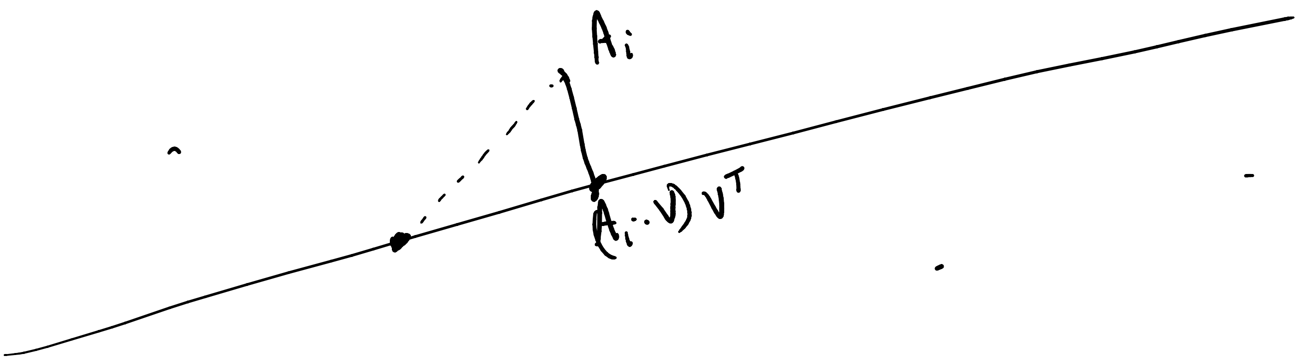
Q. But how to find line joining means?

Best fit line: $v = \underset{\|v\|=1}{\operatorname{argmax}} \|Av\|^2$

maximizes sum of squared projections of rows of A.

For each row A_i

$$\|A_i\|^2 = (A_i v)^2 + \|A_i - (A_i v) v^T\|^2$$



$\|A_i - (A_i v) v^T\|^2$

$$\arg\max \|Av\|^2 = \arg\min \|A - (Av)v^T\|^2$$

least squared error.

$A \in \mathbb{R}^{n \times d}$

u, v : left, right singular vectors of A

$Av = \sigma u \qquad \sigma \geq 0.$

$A^T u = \sigma v$

<u>Lemma</u> $v = \underset{\|v\|=1}{\arg\max} \|Av\|^2$ is a right singular vector of A with largest singular value.

<u>Pf.</u>

$$A^T A v = A^T(\sigma u) = \sigma^2 v$$

v is an eigen vector of $A^T A$.

goes both ways: $A^T A v = \sigma^2 v$

define $u = \frac{1}{\sigma} Av$

Then $A^T u = \frac{1}{\sigma} A^T A v = \sigma v.$

Then $A^T u = \frac{1}{\sigma} A^T A v = \sigma v.$

Now consider $f(v) = \|Av\|^2 = v^T A^T A v.$

$$\nabla_v f(v) = 2A^T A v$$

at any local max/min $\nabla_v f(v) = \lambda v$

$$\Rightarrow \quad A^T A v = \lambda v.$$

Hence the maximizer is an eigenvector.

$$\left.\begin{aligned} &\|Av\|^2 + \lambda(1 - \|v\|^2) \\ &2A^T A v + 2\lambda v = 0 \\ &\|v\|^2 = 1 \end{aligned}\right]$$

SVD.

$v_1 = \arg\max \|Av_1\| \qquad \sigma_1 = \|Av_1\|$

$v_2 = \underset{v_2 \perp v_1}{\arg\max} \|Av_1\|$

$\vdots$

$\|A\ldots\|$

$\vdots$

$$v_k = \underset{v_k \perp v_1 \dots v_{k-1}}{\operatorname{argmax}} \|A v_k\| \qquad \sigma_k .$$

Thm. $V_k = \operatorname{span}\{v_1 \dots v_k\}$ satisfies

$$V_k = \underset{V : \dim(V) = k}{\operatorname{argmin}} \sum_i d(A_{(i)}, V)^2$$

minimizes sum of squared distances among all k-dim subspaces. $v_1, \dots v_k$ are singular vectors with singular values $\sigma_1 \dots \sigma_k \dots$

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

Pf. By induction on k.

V_1 ✓

Suppose true for V_{k-1}.

Consider V_k' OPT k-dim subspace.

... an orthonormal basis

$\dots \; v_k$

Let $w_1 \ldots\ldots w_k$ be an orthonormal basis of V_k' with $w_k \perp V_{k-1}$. ($\exists$ such a w_k).

$\therefore V_k'$ maximizes

$$\|Aw_1\|^2 + \|Aw_2\|^2 + \ldots \|Aw_k\|^2$$

$$\leq \sum_{i=1}^{n} d(A_{(i)}, V_{k-1})^2 + \|Aw_k\|^2$$

$$= \|Av_1\|^2 + \ldots + \|Av_{k-1}\|^2 + \|Av_k\|^2$$

Hence WLOG $V_k' = \text{Span } V_{k-1} \cup \{w_k\}$

But $w_k \perp V_{k-1}$ and must be a maximizer.

i.e. $V_k = V_k'$.

$\left(\sum \sigma u_i v_i^T\right) v_j$ is the same as Av_j

Hence also for any x $(= \sum \alpha_j v_j)$. $\blacksquare$

Back to k-GMMs.

Back to k-GMMs.

Thm. V_k for a mixture of spherical Gaussians $\supseteq \mathrm{Span}\{\mu_1 \dots \mu_k\}$.

Algorithm. - Project sample to top k-dim SVD subspace.

- Cluster according to distances in $\mathbb{R}^k$.

Thm. $|\mu_i - \mu_j| > C \left(\log \frac{m}{\delta} \cdot k\right)^{\frac{1}{4}} \max\{\sigma_i, \sigma_j\}$

suffices!

---

k instead of d.