The Number of Components in Gaussian Mixture Model

The lower bound and rate-optimal estimator for minimax risk estimation

Chi-Ning Chou[1]

[1]Department of Computer Science and Information Engineering
National Taiwan University

July 15, 2015

Outline

Outline

Probability Distribution Family

We consider the following probability distribution family in Gaussian mixture model:

Definition

A Gaussian mixture model family with minimum mass k and minimum central distance δ is

$$D_{k,\delta} = \{g : g = \sum_i \lambda_i f_i, f_i \sim N(\mu_i, 1),$$
$$\sum_i \lambda_i = 1, \lambda_i \geq \frac{1}{k}, |\mu_i - \mu_j| \geq \delta\}$$

- Assume that all the components have the same variance and normalize to 1.

Minimax Risk in $D_{k,\delta}$

As to the error estimation, we choose the minimax scheme:

Definition

The **risk** of an estimator and the **minimax risk** of an estimation on the family $D_{k,\delta}$ is:

$$r_{n,k,\delta}(\hat{S}) := \sup_{\mathbf{X}^n \leftarrow f \in D_{k,\delta}} E[|\hat{S}(\mathbf{X}^n) - S(f)|^2]$$

$$R_{n,k,\delta} := \inf_{\hat{S}} r_{n,k,\delta}(\hat{S})$$

, where n is the number of samples.

Histogram

Definition

$$N_\rho = \{N_s := \sum_j \mathbf{1}_{\{(s-\frac{1}{2})\rho \leq X_j < (s+\frac{1}{2})\rho\}} | s = n\delta, \forall n \in \mathbf{R}\}$$

To construct an estimator, we need to define the **histogram** for convenience.
Intuitively, N_s counts the number of samples around a possible center: $s = n\delta$ of a component with width ρ. Namely, the interval: $[(s - \frac{1}{2})\rho, (s + \frac{1}{2})\rho]$.

Direct Estimator

Thus, we can construct a direct estimator according to the definition of histogram:

Direct Estimator

We construct a direct estimator as

$$\hat{S} := \sum_{s} \mathbf{1}_{\{N_s \geq \frac{n}{k}\Phi(\rho)\}}$$

, where Φ is a scaling function according to the quantization size ρ. Here, we take $\Phi(\rho) = \frac{1}{2}(1 - 2Q(\frac{\rho}{2}))$.

Estimation Error (Risk)
Intuition

There are two kinds of error which depend on whether the sample points lying in the histogram or not:

- False Negative
  - There is actually a component on position s, but there are not enough samples contribute to the histogram. Thus, the estimator doesn't recognize it.
- False Positive
  - There isn't a component on position s, but the samples contribute to N_s are so much that the estimator counts it.

Estimation Error (Risk)
Formally

The error in the scheme is Mean Square Error (MSE). Consider the MSE for a single distribution $f \in D_{k,\delta}$:

$$\begin{aligned}\text{MSE} &= E[|\hat{S}(\mathbf{X}^n) - S(f)|^2] \\ &= E[|\hat{S}(\mathbf{X}^n) - S(f)|]^2 + \text{Var}[|\hat{S}(\mathbf{X}^n) - S(f)|] \\ &= \text{Bias}^2 + \text{Var}\end{aligned}$$

Outline

Bias

The bias of the estimator $\hat{S}$ is:

$$
\begin{aligned}
|E[\hat{S}(P) - S(P)]| &= |\sum_s Pr[N_s \geq \frac{n}{k}\Phi(\rho)] - \mathbf{1}_{\{\exists i \ s.t. \ \mu_i = s\rho\}}| \\
&= \sum_s Pr[N_s < \frac{n}{k}\Phi(\rho)]\mathbf{1}_{\{\exists i, \ \mu_i = s\rho\}} \\
&+ \sum_s Pr[N_s \geq \frac{n}{k}\Phi(\rho)]\mathbf{1}_{\{\forall i, \ \mu_i \neq s\rho\}} \\
&= \sum_s Pr[False\ Negative\ at\ s] \\
&+ \sum_s Pr[False\ Positive\ at\ s]
\end{aligned}
$$

Flase Negative

Reduce to Finite Cases

Observation (From Infinite to Finite)

The false negative probability is:

$$Pr[False\ Negative] = \sum_{s} Pr[False\ Negative\ at\ s]$$
$$= \sum_{i} Pr[False\ Negative\ at\ N_{\mu_i}]$$

That is, we only need to consider finite many cases. Concretely, less than k cases.

False Negative
Single Realization

Observation (Single Realization in N_{μ_i})

The probability of **a single** sample falling inside $[\mu_i - \frac{\rho}{2}, \mu_i + \frac{\rho}{2}]$ is:

$$
\begin{aligned}
P_i &:= Pr[\ A\ single\ realization \in [\mu_i - \frac{\rho}{2}, \mu_i + \frac{\rho}{2}]\ ] \\
&= [1 - 2\lambda_i Q(\frac{\rho}{2})] + \sum_{j \neq i} \frac{\lambda_j}{\sqrt{2\pi}} \int_{\mu_j - \frac{\rho}{2}}^{\mu_j + \frac{\rho}{2}} e^{\frac{-(x-\mu_j)^2}{2}} dx \\
&\geq [1 - 2\lambda_i Q(\frac{\rho}{2})]
\end{aligned}
$$

, where Q, defined as $Q(\alpha) := \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-\frac{t^2}{2}} dt$, is the Q function.

False Negative

n Realizations

Observation (False Negative of n Samples)

Finally, let's consider the false negative probability of n inputs on the ith component. We can observe that the number of samples falling in $[\mu_i - \frac{\rho}{2}, \mu_i + \frac{\rho}{2}]$ is obeying a binomial distribution with n samples and probability P_i^{FN}. That is,

$$Pr[N_{\mu_i} \leq m] = \sum_{j=1}^{m} (\binom{n}{j}) P_i^j (1 - P_i)^{n-j}$$

Thus, the false negative probability of the ith component is:

$$P_i^{FN(n,k,\rho)} := Pr[N_{\mu_i \leq \frac{n}{k}\Phi(\rho)}]$$

False Negative

Upper Bound For False Negative Probability at N_{μ_i}

Proposition (Upper Bound For False Negative Probability at N_{μ_i})

Let P_i^{FN} denotes the false negative probability at N_{μ_i}, then

$$P_i^{FN} \leq \exp(-\frac{n}{2k^2 P_i}\Phi(\rho)^2)$$

False Negative

Upper Bound For False Negative Probability at N_{μ_i} - Proof

Proof.

By the construction of $\hat{S}$,

$$P_i^{FN} = Pr[N_i < \frac{n}{k}\Phi(\rho)]$$

Since

$$\frac{n}{k}\Phi(\rho) = \frac{n(1-2Q(\frac{\rho}{2}))}{2k} \leq \frac{n\lambda_i(1-2Q(\frac{\rho}{2}))}{2} \leq nP_i$$

By Chernoff's inequality [1], when $\frac{n}{k}\Phi(\rho) \leq nP_i$,

False Negative

Upper Bound For False Negative Probability at N_{μ_i} - Proof

Proof.

$$
\begin{aligned}
Pr[N_i < \frac{n}{k}\Phi(\rho)] &\leq \exp(-\frac{1}{2P_i}\frac{(nP_i - \frac{n}{k}\Phi(\rho))^2}{n}) \\
&= \exp(-\frac{n}{2k^2P_i}(kP_i - \Phi(\rho))^2)
\end{aligned}
$$

Moreover,

$$
2\Phi(\rho) = (1 - 2Q(\frac{\rho}{2})) \leq k\lambda_i(1 - 2Q(\frac{\rho}{2})) \leq kP_i
$$

False Negative

Upper Bound For False Negative Probability at N_{μ_i} - Proof

Proof.

Thus,

$$(kP_i - \Phi(\rho))^2 \geq (2\Phi(\rho) - \Phi(\rho))^2 = \Phi(\rho)^2$$

That is,

$$\begin{aligned} P_i^{FN} &\leq \exp(-\frac{n}{2k^2 P_i}(kP_i - \Phi(\rho))^2) \\ &\leq \exp(-\frac{n}{2k^2 P_i}\Phi(\rho)^2) \end{aligned}$$

$\square$

False Negative

Upper Bound For False Negative Probability

Proposition (Upper Bound For False Negative Probability)

Let P^{FN} denotes the false negative probability, then

$$P^{FN} \leq k\exp(-\frac{n}{2k}\Phi(\rho)^2)$$

False Negative

Upper Bound For False Negative Probability - Proof

Proof.

From the previous results,

$$P^{FN} = \sum_i P_i^{FN} \leq \sum_i \exp(-\frac{n}{2k^2 P_i}\Phi(\rho)^2)$$

Since, $\sum_i P_i \leq 1$ and $P_i \geq \frac{1}{k}\ \forall i$.

$$P^{FN} \leq \sum_i \exp(-\frac{n}{2k^2 P_i}\Phi(\rho)^2) \leq k \exp(-\frac{n}{2k}\Phi(\rho)^2)$$

$\square$

Outline

1. Problem Setting
   - Previous Work
2. Analysis On Estimator
   - Bias
   - Variance
3. Appendix
   - Inequalities of Q Function

Outline

The Gaussian Tail Inequality - 1

Theorem (The Gaussian Tail Inequality - 1)

If $X \sim N(0,1)$, then

$$\frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{\alpha^2}{2}} (1 - \frac{1}{\alpha^2}) < Q(\alpha) < \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{\alpha^2}{2}}$$

Corollary (False Negative Version)

If $X \sim N(0,1)$, then

$$4\lambda_i [\frac{1}{\sqrt{2\pi}\rho} e^{-\frac{\rho^2}{8}} (1 - \frac{4}{\rho^2})] < Pr[FN_i]$$

The Gaussian Tail Inequality - 2

Theorem (The Gaussian Tail Inequality - 2)

If $X \sim N(0,1)$, then

$$\frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{\alpha^2}{2}} \left(\frac{\alpha^2}{1+\alpha^2}\right) < Q(\alpha)$$

Corollary (False Negative Version)

If $X \sim N(0,1)$, then

$$4\lambda_i\left[\frac{1}{\sqrt{2\pi}\rho} e^{-\frac{\rho^2}{8}} \left(\frac{\rho^2}{4+\rho^2}\right)\right] < Pr[FN_i]$$

The Gaussian Tail Inequality - 3

Theorem (The Gaussian Tail Inequality - 3)

If $X \sim N(0,1)$, then

$$Q(\alpha) \leq \frac{1}{2}e^{-\frac{\alpha^2}{2}}$$

Corollary (False Negative Version)

This one doesn't give a lower bound for false negative.

The Gaussian Central Inequality

This inequality gives a tight bound near the center of both Q function and error function.

Theorem (The Gaussian Central Inequality)

If $X \sim N(0,1)$, then

$$max\{1-\frac{1}{2}e^{\sqrt{\frac{2}{\pi}}\alpha},0\} \leq Q(\alpha) \leq min\{\frac{1}{2}e^{-\sqrt{\frac{2}{\pi}}\alpha},1\}$$

Corollary (False Negative Version)

If $X \sim N(0,1)$, then

$$max\{2\lambda_i-\lambda_i e^{\frac{\alpha}{\sqrt{2\pi}}},0\} \leq Pr[FN_i]$$

For Further Reading I

Wikipedia.
Tail Bounds of Binomial Distribution

S. Someone.
On this and that.
Journal of This and That, 2(1):50–100, 2000.