Learning mixtures of spherical Gaussians: moment methods and spectral decompositions

Daniel Hsu and Sham Kakade

Arthur Gretton's notes

September 22, 2013

Problem setup

A very basic topic model with k topics.
Probability of a topic

$$P(h = j) = w_j, \quad j \in [k],$$

where $[k] := \{1, \ldots, k\}$, h is document topic.

Problem setup

A very basic topic model with k topics.
Probability of a topic

$$P(h = j) = w_j, \quad j \in [k],$$

where $[k] := \{1, \ldots, k\}$, h is document topic.
Documents have length ℓ, $t \in [\ell]$.
If word t is the ith word, then

$$x_t = e_i$$

where e_i is the unit vector in dimension i.
Vocabulary is of size d.

Problem setup

Probabilty vector for topic j is $\mu_j \in \Delta^{d-1}$.
Expected *conditional* probability of t th word in a document, given *topic* j:

$$E(x_t | h = j) = \sum_{i=1}^{d} [\mu_j]_i e_i = \mu_j,$$

$$M := \begin{bmatrix} \mu_1 & \dots & \mu_k \end{bmatrix}.$$

Problem setup

Probabilty vector for topic j is $\mu_j \in \Delta^{d-1}$.
Expected *conditional* probability of t th word in a document, given *topic* j:

$$E(x_t | h = j) = \sum_{i=1}^{d} [\mu_j]_i e_i = \mu_j,$$

$$M := \begin{bmatrix} \mu_1 & \dots & \mu_k \end{bmatrix}.$$

Moments:

$$M_2 = E[x_1 \otimes x_2] = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i = M \mathrm{diag}(w) M^\top$$

$$M_3 = E[x_1 \otimes x_2 \otimes x_3] = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

Question: can we learn w_i, μ_i from empirical moments?

Just a moment?

Do we need M_3?
Recall:

$$M_2 = M\mathrm{diag}(w)M^\top \underset{?}{=} \tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top$$

for some new $\tilde{M}$, $\tilde{w}$.
Define

$$\tilde{M} = MQ^{-1} \qquad \tilde{w} = Qw.$$

What properties must $Q \in \Re^{k\times k}$ satisfy?

Just a moment?

1. Q should "look like" a matrix of conditional probabilities; i.e., each column should sum to 1:
$$1^\top Q = 1^\top.$$
2. MQ^{-1} and Qw should have non-negative entries. So should M_2,
$$\tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top = MQ^{-1}\mathrm{diag}(Qw)\left(Q^{-1}\right)^\top M^\top =: M_2$$
3. For the proof to work: $Q\mathrm{diag}(w)Q^\top$ should have diagonal entries.

Just a moment?

Proof:

$$M\mathrm{diag}(w)M^\top = \tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top.$$

Substituting definitions for $\tilde{M}$ and $\tilde{w}$, we get

$$\tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top = MQ^{-1}\mathrm{diag}(Qw)\left(Q^{-1}\right)^\top M^\top$$

Just a moment?

Proof:

$$M\mathrm{diag}(w)M^\top = \tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top.$$

Substituting definitions for $\tilde{M}$ and $\tilde{w}$, we get

$$\tilde{M}\mathrm{diag}(\tilde{w})\tilde{M}^\top = MQ^{-1}\mathrm{diag}(Qw)\left(Q^{-1}\right)^\top M^\top$$

Need to show

$$Q^{-1}\,\mathrm{diag}(Qw)\left(Q^{-1}\right)^\top = \mathrm{diag}(w)$$

under assumptions.

Just a moment?

$$[\mathrm{diag}(Qw)]_{ii} = \sum_{\ell} Q_{i\ell} w_\ell$$

$$\underset{\text{(a)}}{=} \sum_{\ell} \underbrace{\left(\sum_{j=1}^{k} Q_{j\ell} \right)}_{=1} Q_{i\ell} w_\ell$$

Just a moment?

$$
\begin{aligned}
[\mathrm{diag}(Qw)]_{ii} &= \sum_{\ell} Q_{i\ell} w_{\ell} \\
&\underset{\text{(a)}}{=} \sum_{\ell} \underbrace{\left(\sum_{j=1}^{k} Q_{j\ell} \right)}_{=1} Q_{i\ell} w_{\ell} \\
&= \sum_{j} \sum_{\ell} Q_{j\ell} w_{\ell} Q_{i\ell}
\end{aligned}
$$

Just a moment?

$$
\begin{aligned}
[\mathrm{diag}(Qw)]_{ii} &= \sum_{\ell} Q_{i\ell} w_\ell \\
&\underset{(a)}{=} \sum_{\ell} \underbrace{\left(\sum_{j=1}^{k} Q_{j\ell} \right)}_{=1} Q_{i\ell} w_\ell \\
&= \sum_{j} \sum_{\ell} Q_{j\ell} w_\ell Q_{i\ell} \\
&= \left(1^\top Q \mathrm{diag}(w) Q^\top \right)_i \\
&= \left(Q \mathrm{diag}(w) Q^\top \right)_{ii},
\end{aligned}
$$