Bhattacharyya clustering with applications to mixture simplifications
ICPR 2010, Istanbul, Turkey

Frank Nielsen\textsuperscript{1,2} Sylvain Boltz\textsuperscript{1} Olivier Schwander\textsuperscript{1,3}

\textsuperscript{1}\textsuperscript{1}\textsuperscript{1}École Polytechnique, France
\textsuperscript{2}Sony Computer Science Laboratories, Japan
\textsuperscript{3}ÉNS Cachan, France

August, 24 2010
Mean
Definition
Burbea-Rao divergences
Burbea-Rao centroid

Exponential Family
Definition
Bhattacharyya distance
Closed-form formula

Application
Statistical mixtures
Mixture simplification
Introduction

Bhattacharyya distance

- Widely used to compare probability density functions
- Good statistical properties, related to Fisher information
- Measures the overlap between two distributions

Bhattacharyya coefficient

\[ B_c(p, q) = \int \sqrt{p(x)q(x)} \, dx \leq 1 \]

Bhattacharyya distance

\[ B(p, q) = -\log B_c(p, q) \geq 0 \]
Contributions

Drawbacks
- Few closed-form formula are known
- Centroid estimation only for univariate Gaussian, without guarantees

Results
- Bhattacharyya between exponential families, using Burbea-Rao divergences
- Efficient scheme for centroid
- Application to simplification of Gaussian mixtures
What is a mean?

Euclidean geometry

- Given a set of $n$ points $\{p_i\}$,
- the center of mass (a.k.a. center of gravity) is

$$c = \frac{1}{n} \sum_{i} p_i$$

Unique minimizer of average squared Euclidean distance

$$c = \arg \min_p \sum_{i} \|p - p_i\|^2$$

Definitions

- By axiomatization
- By optimization
Axiomatization

Axioms for a mean function $M(x_1, x_2)$

- Reflexivity: $M(x, x) = x$
- Symmetry: $M(x_1, x_2) = M(x_2, x_1)$
- Continuity: $M(\cdot, \cdot)$ continuous
- Strict monotonicity: $M(x_1, x_2) < M(x'_1, x_2)$ for $x_1 < x'_1$
- Anonymity:
  \[
  M(M(x_{11}, x_{12}), M(x_{21}, x_{22})) = M(M(x_{11}, x_{21}), M(x_{12}, x_{22}))
  \]

Yields to a unique family

\[
M(x_1, x_2) = f^{-1} \left( \frac{f(x_1) + f(x_2)}{2} \right)
\]

with $f$ continuous, strictly monotonous and increasing function
Examples and $f$-representation

Some $f$-means

- Arithmetic mean: $\frac{x_1 + x_2}{2}$ with $f(x) = x$
- Geometric mean: $\sqrt{x_1 x_2}$ with $f(x) = \log x$
- Harmonic mean: $\frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$ with $f(x) = \frac{1}{x}$

Arithmetic mean on the $f$-representation

- $y = f(x)$
- $f(\bar{x}) = \frac{1}{n} \sum_i f(x_i)$
- $\bar{y} = \frac{1}{n} \sum_i y_i$
Optimization

Problem

\[
\min_x \sum_i \omega_i d(x, p_i) = \min_x L(x; (\{x_i\}, \{\omega_i\}), d)
\]

Entropic mean (Ben-Tal et al., 1989)

- \(d(p, q) = I_f(p, q) = pf\left(\frac{q}{p}\right)\) (Csiszar \(f\)-divergence)
- \(f\) is a strictly convex differentiable function with \(f(1) = 0\) and \(f'(1) = 0\)

Some entropic means

- Arithmetic mean: \(f(x) = -\log x + x - 1\)
- Geometric mean: \(f(x) = x \log x - x + 1\)
- Harmonic mean: \(f(x) = (x - 1)^2\)
Bregman means

Bregman divergence

- $B_F(p, q) = F(p) - F(q) + \langle p - q \mid \nabla F(q) \rangle$
- $F$ is a strictly convex and differentiable function

Convex problem

- unique minimizer
- $c = \nabla F^{-1} (\sum_i \omega_i \nabla F(x_i))$

Since $B_F$ is not symmetrical, there is another centroid

- Left-sided one: $\min_x \sum_i \omega_i B_F(x, p_i)$
- Right-sided one: $\min_x \sum_i \omega_i B_F(p_i, x)$
Burbea-Rao divergence

Based on Jensen inequality for a convex function $F$

$$BR_F(p, q) = \frac{F(p) + F(q)}{2} - F\left(\frac{p + q}{2}\right) \geq 0$$

Special case: Jensen-Shannon divergence

- $JS(p, q) = KL(p, \frac{p+q}{2}) + KL(q, \frac{p+q}{2})$
- $JS(p, q) = H\left(\frac{p+q}{2}\right) - \frac{H(p) + H(q)}{2} \geq 0$
- $H(x) = -F(x) = -x \log x$ (Shannon entropy)
Symmetrizing Bregman divergences

Jeffreys-Bregman divergence

\[ S_F(p, q) = \frac{1}{2} (B_F(p, q) + B_F(q, p)) \]
\[ = \frac{1}{2} \langle p - q | \nabla F(p) - \nabla F(q) \rangle \]

Jensen-Bregman divergence

\[ J_F(p, q) = \frac{1}{2} \left( B_F(p, \frac{p + q}{2}) + B_F(q, \frac{p + q}{2}) \right) \]
\[ = \frac{F(p) + F(q)}{2} - F \left( \frac{p + q}{2} \right) \]
\[ = BR_F(p, q) \]
Burbea-Rao centroid

Optimization problem

\[ c = \arg \min_x \sum_i \omega_i BR_F(x, p_i) = \arg \min L(x) \]

\[ L(x) \equiv \frac{1}{2} F(x) - \sum_i \omega_i F\left(\frac{c + p_i}{2}\right) \]

ConCave Convex Procedure (CCCP, NIPS2001)

- iterative scheme
- \[ \nabla L_{\text{convex}}(x^{(k+1)}) = \nabla L_{\text{concave}}(x^{(k)}) \]
- converges to a local minimum
ConCave Convex Procedure

Possible decomposition for function with bounded Hessian
Iterative algorithm for Burbea-Rao centroids

Initialization
\( x^{(0)} \): center of mass (Bregman right-sided centroid), or symmetrized KL divergence

Iteration

\[
\nabla F(x^{(k+1)}) = \sum_i \omega_i \nabla F \left( \frac{x^{(t)} + p_i}{2} \right)
\]

Centroid

\[
\[x^{(t+1)} = \nabla F^{-1} \left( \sum_i \omega_i \nabla F \left( \frac{x^{(t)} + p_i}{2} \right) \right)
\]
Exponential family

Definition

\[ p(x; \lambda) = p_F(x; \theta) = \exp (\langle t(x) | \theta \rangle - F(\theta) + k(x)) \]

- \( \lambda \) source parameter
- \( \theta \) natural parameter
- \( F(\theta) \) log-normalizer
- \( k(x) \) carrier measure
Example

Poisson distribution

\[ p(x; \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda) \]

- \( t(x) = x \)
- \( \theta = \log \lambda \)
- \( F(\theta) = \exp(\theta) \)
Multivariate normal distribution

Gaussian

\[
p(x; \mu, \Sigma) = \frac{1}{2\pi^{\frac{d}{2}} \sqrt{\det \Sigma}} \exp \left( -\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right)
\]

Exponential family

- \( \theta = (\theta_1, \theta_2) = (\Sigma^{-1} \mu, \frac{1}{2} \Sigma^{-1}) \)
- \( F(\theta) = \frac{1}{4} \text{tr} (\theta_1^{-1} \theta_2 \theta_2^T) - \frac{1}{2} \log \det \theta_1 + \frac{d}{2} \log \pi \)
- \( t(x) = (x, -x^T x) \)
- \( k(x) = 0 \)

Composite vector-matrix inner product

\[
\langle \theta, \theta' \rangle = \theta_1^T \theta_1' + \text{tr}(\theta_2^T \theta_2')
\]
Bhattacharyya distance

Bhattacharyya coefficient

- Amount of overlap between distributions
- \( B_c(p, q) = \int \sqrt{p(x)q(x)} \, dx \)

Bhattacharyya distance

- \( B(p, q) = -\log B_c(p, q) \)

Metrization

- Hellinger-Matusita metric
- \( H(p, q) = \sqrt{1 - B(p, q)} \)
- Gives the same Voronoi diagram
Closed-form formula

\[ B_c(p, q) = \int \sqrt{p(x)q(x)} \, dx \]
\[ = \int \exp \left( \langle t(x), \frac{\theta_p + \theta_q}{2} \rangle - \frac{F(\theta_p + \theta_q)}{2} + k(x) \right) \, dx \]
\[ = \exp \left( F \left( \frac{\theta_p + \theta_q}{2} \right) - \frac{F(\theta_p) + F(\theta_q)}{2} \right) > 0 \]

\[ B(p, q) = -\log B_c(p, q) = BR_F(\theta_p, \theta_q) \geq 0 \]

Equivalence

- Bhattacharyya between two members of the same EF
- Burbea-Rao between natural parameters using log-normalizer
## Examples

<table>
<thead>
<tr>
<th>Exponential family</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multinomial</td>
<td>$-\ln \sum_{i=1}^{d} \sqrt{p_i q_i}$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\frac{1}{2} \left( \sqrt{\mu_p} - \sqrt{\mu_q} \right)^2$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\frac{1}{4} \frac{(\mu_p - \mu_q)^2}{\sigma_p^2 + \sigma_q^2} + \frac{1}{2} \ln \frac{\sigma_p^2 + \sigma_q^2}{2 \sigma_p \sigma_q}$</td>
</tr>
<tr>
<td>Multivariate Gaussian</td>
<td>$\frac{1}{8} (\mu_p - \mu_q)^t \left( \frac{\Sigma_p + \Sigma_q}{2} \right)^{-1} (\mu_p - \mu_q) + \frac{1}{2} \ln \frac{\det \frac{\Sigma_p + \Sigma_q}{2}}{\det \Sigma_p \det \Sigma_q}$</td>
</tr>
</tbody>
</table>
Gaussian Mixture Models

Mixture

- $\Pr(X = x) = \sum_i \omega_i \Pr(X = x | \mu_i, \Sigma_i)$
- each $\Pr(X = x | \mu_i, \Sigma_i)$ is a multivariate normal distribution

Soft Clustering

Expectation-Maximization algorithm, equivalent to soft Bregman clustering
Statistical images

http://www.informationgeometry.org/MEF/

RGBxy representation: 5D point set
Mixture simplification

Initialization

- Mixture of Gaussians, with Bregman soft clustering (≡ EM)

Simplification

- $k$-means using Bhattacharyya distance and centroids

Different $k$

- Hierarchical clustering
Hierarchical clustering

(a) source

(b) $k = 48$

(c) $k = 16$
Conclusion

Results

- Symmetrizing Bregman yields Burbea-Rao divergences
- Bhattacharyya between exponential families yields Burbea-Rao
- Closed-form formula for Bhattacharyya between EF
- Efficient scheme for BR centroid using CCCP

Applications

- Simplification of Gaussian Mixture Models
- Hierarchical Clustering
References


www.informationgeometry.org