

Taxonomy of principal distances and divergences



Sony CSL

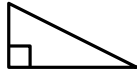


Bolyai (1802-1860)



Lobachevsky (1792-1856)

Euclidean geometry



Hyperbolic/spherical geometry



Statistical geometry

Additive entropy

cross-entropy
conditional entropy
mutual information
(chain rules)

Euclidean distance
 $d_2(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_i (p_i - q_i)^2}$ (Pythagoras' theorem circa 500 BC)



Hamming distance
 $(|\{i : p_i \neq q_i\}|)$

Manhattan distance
 $d_1(\mathbf{p}, \mathbf{q}) = \sum_i |p_i - q_i|$
(city block-taxi cab)

Minkowski distance (L_k -norm)
 $d_k(\mathbf{p}, \mathbf{q}) = \sqrt[k]{\sum_i |p_i - q_i|^k}$
(H. Minkowski 1864-1909)



Lévy-Prokhorov distance
 $LP_\rho(p, q) = \inf_{\epsilon > 0} \{p(A) \leq q(A^\epsilon) + \epsilon \forall A \in \mathcal{B}(\mathcal{X})\}$
 $A^\epsilon = \{y \in \mathcal{X}, \exists x \in A : \rho(x, y) < \epsilon\}$

Quadratic distance
 $d_Q = \sqrt{(\mathbf{p} - \mathbf{q})^T \mathbf{Q} (\mathbf{p} - \mathbf{q})}$

Riemannian geometry



Riemannian metric tensor
 $\int \sqrt{g_{ij} \frac{dx_i}{ds} \frac{dx_j}{ds}} ds$
(B. Riemann 1826-1866.)



Fisher-Rao distance:
 $ds^2 = g_{ij} d\theta^i d\theta^j = d\theta^T I(\theta) d\theta$
 $\rho_{FR}(p, q) = \min_\gamma \int_0^1 \sqrt{\dot{\gamma}(t)^T I(\theta) \dot{\gamma}(t)} dt$

Affine differential geometry

Logarithmic divergence

$L_{G, \alpha}(\theta_1 : \theta_2) = \frac{1}{\alpha} \log(1 + \alpha \nabla G(\theta_2)^T (\theta_1 - \theta_2)) + G(\theta_2) - G(\theta_1)$

$\alpha \rightarrow 0, F = -G$

Bregman divergences (1967):

$B_F(\theta_1 || \theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2)$

Fisher information (local entropy)

$I(\theta) = E[(\frac{\partial}{\partial \theta} \ln p(X|\theta))^2]$
(R. A. Fisher 1890-1962)



Finsler metric tensor
 $g_{ij} = \frac{1}{2} \partial^2 \frac{F^2(x, y)}{\partial y^i \partial y^j}$

Aitchison distance
Probability simplex

Hilbert
log-ratio metric



Chernoff divergence (1952)
 $C_\alpha(p||q) = -\ln \int p^\alpha q^{1-\alpha} d\mu$
 $C(p, q) = \max_{\alpha \in (0, 1)} C_\alpha(p||q)$



Rényi divergence (1961)
 $H_\alpha = \frac{1}{\alpha(1-\alpha)} \log \int f^\alpha d\mu$
 $R_\alpha(p||q) = \frac{1}{\alpha(1-\alpha)} \ln \int p^\alpha q^{1-\alpha} d\mu$
(additive entropy)



Csiszár' f -divergence
 $D_f(p||q) = \int p f(\frac{p}{q}) d\mu$
(Ali& Silvey 1966, Csizsár 1967)

Dual div. (Legendre) $D_{F^*}(\nabla F(\theta_1) || \nabla F(\theta_2)) = D_F(\theta_2 || \theta_1)$
Dual div. *-conjugate ($F^*(y) = y f(1/y)$)
 $D_{F^*}(p||q) = D_f(q||p)$

Information geometries

Amari α -divergence (1985)
 $f_\alpha(x) = \begin{cases} x \log x & \alpha = 1 \\ -\log x & \alpha = -1 \\ \frac{4}{1-\alpha^2} (1-x)^{\frac{1+\alpha}{2}} & -1 < \alpha < 1 \end{cases}$

Quantum & matrix geometry

Fröbenius & Hilbert-Schmidt norm

Quantum entropy
 $S(\rho) = -k \text{Tr}(\rho \log \rho)$
(Von Neumann 1927)



Quantum f -divergences
(Dénés Petz)

Von Neumann divergence
 $D(\mathbf{P}||\mathbf{Q}) = \text{Tr}(\mathbf{P}(\log \mathbf{P} - \log \mathbf{Q}) - \mathbf{P} + \mathbf{Q})$



Stein discrepancies

Integral probability metrics
IPMs

MMD
Maximum Mean Discrepancy

Gromov-Hausdorff distance

(between compact metric spaces)
 $d_{GH}(X, Y) = \inf_{\phi_X: X \rightarrow Z, \phi_Y: Y \rightarrow Z} \{\rho_H^Z(\phi_X(X), \phi_Y(Y))\}$
 ϕ_X, ϕ_Y : isometric embeddings

Sinkhorn divergence (h -regularized OT)

Optimal transport geometry

©2022 Frank Nielsen

Earth mover distance
(EMD 1998)
 $\rho = L_1$

Wasserstein distances

$W_{\alpha, \rho}(p, q) = (\inf_{\gamma \in \Gamma(p, q)} \int \rho(p, q)^\alpha d\gamma(x, y))^{1/\alpha}$



Tsallis entropy (1998)
(Non-additive entropy)
 $T_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} (\int p^\alpha d\mu - 1)$
 $T_\alpha(p||q) = \frac{1}{1-\alpha} (1 - \int \frac{p^\alpha}{q^{1-\alpha}} d\mu)$

Non-additive entropy

Sharma-Mittal entropies
 $h_{\alpha, \beta}(p) = \frac{1}{1-\beta} \left((\int p^\alpha d\mu)^{\frac{1-\beta}{1-\alpha}} - 1 \right)$

Generalized Pythagoras' theorem

(Generalized projection)

Bregman-Csiszár divergence (1991)

$F_\alpha(x) = \begin{cases} x - \log x - 1 & \alpha = 0 \\ x \log x - x + 1 & \alpha = 1 \\ \frac{1}{\alpha(1-\alpha)} (-x^\alpha + \alpha x - \alpha + 1) & 0 < \alpha < 1 \end{cases}$

Itakura-Saito divergence

$IS(\mathbf{p}||\mathbf{q}) = \sum_i (\frac{p_i}{q_i} - \log \frac{p_i}{q_i} - 1)$
(Burg entropy)

Generalized
 f -means
duality...

$\beta = 1$

Generalized
 f -means
duality...

Generalized
 f -means
duality...