Distances & Divergences — Quick-Reference Table

Grouped families, defining equations, and typical applications. Generator convention: f-divergence $D_f(p\|q) = \int p\, f(q/p)\, d\mu$; Bregman B_F from convex F. $\|$ marks asymmetric (divergence) arguments; argument-order conventions vary by source.

Group	Distance / divergence	Equation	Application — where to use
Vector & metric distances	Euclidean (L_2)	$d_2 = \sqrt{\sum_i (p_i - q_i)^2}$	k-means, k-NN, least squares
	Manhattan (L_1)	$d_1 = \sum_i \lvert p_i - q_i \rvert$	Grid/route cost; robust to outliers; LASSO
	Minkowski (L_k)	$d_k = \big(\sum_i \lvert p_i - q_i \rvert^k\big)^{1/k}$	Tunable; unifies L_1, L_2, L_∞
	Chebyshev (L_∞)	$d_\infty = \max_i \lvert p_i - q_i \rvert$	Max-coordinate gap; chessboard moves
	Mahalanobis	$\sqrt{(p-q)^\top \Sigma^{-1}(p-q)}$	Scale/correlation-aware; outliers, classification
	Quadratic	$\sqrt{(p-q)^\top Q\,(p-q)}$	Feature-weighted; colour histograms
	Hamming	$\big\lvert \{i : p_i \neq q_i\} \big\rvert$	Error-correcting codes, DNA, bit strings
Riemannian & information geometry	Fisher information	$\mathbf{I}(\theta) = \mathbb{E}[(\partial_\theta \ln p)(\partial_\theta \ln p)^\top]$	Natural metric on models; Cramér–Rao, natural gradient
	Fisher–Rao distance	$\rho = \min_\gamma \int_0^1 \sqrt{\dot\gamma^\top \mathbf{I}\,\dot\gamma}\, dt$	Intrinsic geodesic distance between distributions
	Riemannian geodesic	$L = \int \sqrt{g_{ij}\, \dot x^i \dot x^j}\, dt$	Shortest path on curved manifolds; shape spaces
	Finsler metric	$g_{ij}(x,y) = \frac{1}{2}\, \partial^2 F^2 / \partial y^i \partial y^j$	Direction-dependent (anisotropic) norms
f-divergences (Csiszár / Ali–Silvey)	*general template*	$D_f(p\|q) = \int p\, f(q/p)\, d\mu$	Master family — pick the convex generator f
	Kullback–Leibler	$\int p \log \frac{p}{q}\, d\mu \quad (f{=}-\log t)$	MLE, cross-entropy loss, variational inference
	Reverse KL	$\int q \log \frac{q}{p}\, d\mu \quad (f{=}t \log t)$	Mode-seeking VI / expectation propagation
	Pearson χ^2	$\int \frac{(q-p)^2}{p}\, d\mu \quad (f{=}(t{-}1)^2)$	Goodness-of-fit; local KL approximation
	Neyman χ^2	$\int \frac{(p-q)^2}{q}\, d\mu \quad (f{=}(1{-}t)^2/t)$	Reverse Pearson; importance-sampling variance
	Hellinger	$\sqrt{\frac{1}{2}\int (\sqrt{p} - \sqrt{q})^2\, d\mu}$	Symmetric bounded *metric*; robust statistics
	Total variation	$\frac{1}{2}\int \lvert p - q \rvert\, d\mu$	Statistical distinguishability; coupling
	Amari α-divergence	$f_\alpha(t) = \frac{4}{1-\alpha^2}\big(1 - t^{\frac{1+\alpha}{2}}\big)$	One dial: KL ($\alpha{=}1$), Hellinger ($\alpha{=}0$)
Bregman divergences	*general template*	$B_F(x\|y) = F(x) - F(y) - \langle x - y, \nabla F(y) \rangle$	Master family — pick the convex potential F
	Squared Euclidean	$\|x - y\|^2$	Centroid clustering (k-means)
	Generalized KL	$\sum_i \big(x_i \log \frac{x_i}{y_i} - x_i + y_i\big)$	Unnormalized KL; NMF, Poisson data
	Itakura–Saito	$\sum_i \big(\frac{p_i}{q_i} - \log \frac{p_i}{q_i} - 1\big)$	Audio/spectral distortion; NMF
	Mahalanobis ($F{=}x^\top A x$)	$(x-y)^\top A\,(x-y)$	Feature-weighted; metric learning
	Log-Det	$\langle P, Q^{-1} \rangle - \log\det(PQ^{-1}) - n$	SPD matrices; metric learning (ITML)
Matrix Bregman divergences	Squared Frobenius	$\|X - Y\|_F^2 = \mathrm{Tr}\big((X-Y)^2\big)$	Baseline matrix distance (matrix Mahalanobis)
	Von Neumann	$\mathrm{Tr}\big(X(\log X - \log Y) - X + Y\big)$	Density/covariance matrices; quantum rel. entropy
	Log-Det / Burg	$\mathrm{Tr}(XY^{-1}) - \log\det(XY^{-1}) - n$	Covariance comparison; portfolio selection
	Bregman–Schatten p	$\frac{1}{2}\,\mathrm{Tr}\big(X^{2p} - 2XY^{p-1} + (p{-}1)Y^p\big)$	Tunable spectral family ($p > 1$)

Group	Distance / divergence	Equation	Application — where to use
Overlap / α-power family	Bhattacharyya	$-\log\int\sqrt{pq}\ d\mu$	Class separability; object tracking
	Chernoff information	$\max_{\alpha\in(0,1)} -\log\int p^{\alpha}q^{1-\alpha}d\mu$	Error exponent in hypothesis testing
	Rényi divergence	$\frac{1}{\alpha-1}\log\int p^{\alpha}q^{1-\alpha}d\mu$	Differential privacy; information theory
Symmetrized & Jensen-type	Jeffreys	$\mathrm{KL}(p\\|q)+\mathrm{KL}(q\\|p)$	Symmetric KL when direction is arbitrary
	Jensen–Shannon	$\frac{1}{2}\mathrm{KL}(p\\|m)+\frac{1}{2}\mathrm{KL}(q\\|m),\ m=\frac{p+q}{2}$	Bounded; $\sqrt{\cdot}$ is a metric; GANs
	Burbea–Rao / Jensen	$\frac{F(p)+F(q)}{2}-F\big(\frac{p+q}{2}\big)$	Jensen gap; JS is the Shannon-entropy case
Entropies (functionals behind divergences)	Shannon / Boltzmann	$H=-\int p\log p\ d\mu$	Information content, source coding
	Rényi	$H_\alpha=\frac{1}{1-\alpha}\log\int p^{\alpha}d\mu$	Min-/collision-entropy; cryptography
	Tsallis (non-additive)	$T_\alpha=\frac{1}{1-\alpha}\big(\int p^{\alpha}d\mu-1\big)$	Non-extensive (long-range) systems
	Sharma–Mittal	$\frac{1}{1-\beta}\big((\int p^{\alpha}d\mu)^{\frac{1-\beta}{1-\alpha}}-1\big)$	Two-parameter unifier (Shannon/Rényi/Tsallis)
	Von Neumann (quantum)	$S(\rho)=-\mathrm{Tr}(\rho\log\rho)$	Quantum entropy; entanglement
Set & metric-space distances	Hausdorff	$\max\{\sup_x\rho(x,Y),\ \sup_y\rho(X,y)\}$	Set/shape matching; image comparison
	Gromov–Hausdorff	$\inf_{\phi_X,\phi_Y}\rho_H^Z(\phi_X X,\phi_Y Y)$	Compare spaces up to isometry; manifold learning
Optimal transport & IPMs	Wasserstein / EMD	$\big(\inf_{\gamma\in\Gamma}\int\rho(x,y)^{\alpha}d\gamma\big)^{1/\alpha}$	Optimal transport; WGAN, retrieval; EMD$=W_1$
	Max Mean Discrepancy	$\sup_{\\|f\\|_{\mathcal{H}}\le 1}\lvert\mathbb{E}_p f-\mathbb{E}_q f\rvert$	Kernel two-sample tests; generative models
	Stein discrepancy	$\sup_{f\in\mathcal{F}}\big\lvert\mathbb{E}_p[\,f\,\nabla\log q+\nabla f\,]\big\rvert$	Sampler/MCMC diagnostics; SVGD (score only)
	Kolmogorov–Smirnov	$\sup_x\lvert F_p(x)-F_q(x)\rvert$	Non-parametric goodness-of-fit
	Lévy–Prokhorov	$\inf\{\varepsilon: p(A)\le q(A^{\varepsilon})+\varepsilon\ \forall A\}$	Metrizes weak convergence (in distribution)
Quantum geometry	Von Neumann divergence	$\mathrm{Tr}\big(P(\log P-\log Q)-P+Q\big)$	Quantum KL; distinguishing quantum states

Nesting at a glance: $L_1\subset L_2\subset L_\infty$ are Minkowski cases; Mahalanobis is Quadratic with $Q=\Sigma^{-1}$. KL is the *only* divergence that is both an f-divergence *and* a Bregman divergence; KL/reverse-KL are the $\alpha=\mp 1$ α-divergences. Bhattacharyya/Chernoff/Rényi all come from the affinity $\int p^{\alpha}q^{1-\alpha}d\mu$. Matrix Bregman divergences (Frobenius, von Neumann, Log-Det, Schatten-p) lift Bregman to matrices via a spectral seed and the trace inner product. Wasserstein, MMD, Stein, KS, Lévy–Prokhorov are IPMs. Reorganized from Frank Nielsen's taxonomy chart (`franknielsen.github.io/Divergence`).