ENM 531: Data-driven Modeling and Probabilistic Scientific Computing

Lecture #2: Primer on Probability and Statistics

Paris Perdikaris
January 19, 2023

Penn
University *of* Pennsylvania

Basic rules of probability

Sum rule

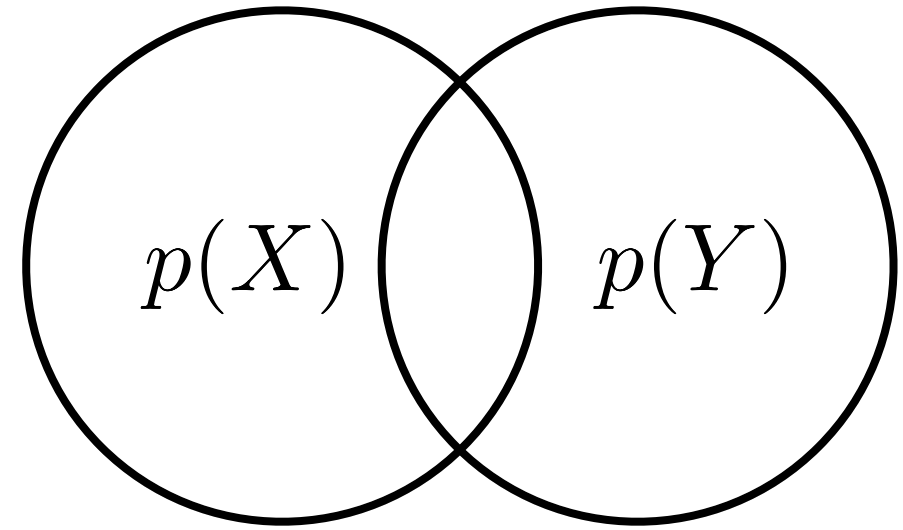
$$p(X) = \sum_Y p(X, Y)$$

Product rule

$$p(X, Y) = p(Y|X)p(X)$$

Bayes rule

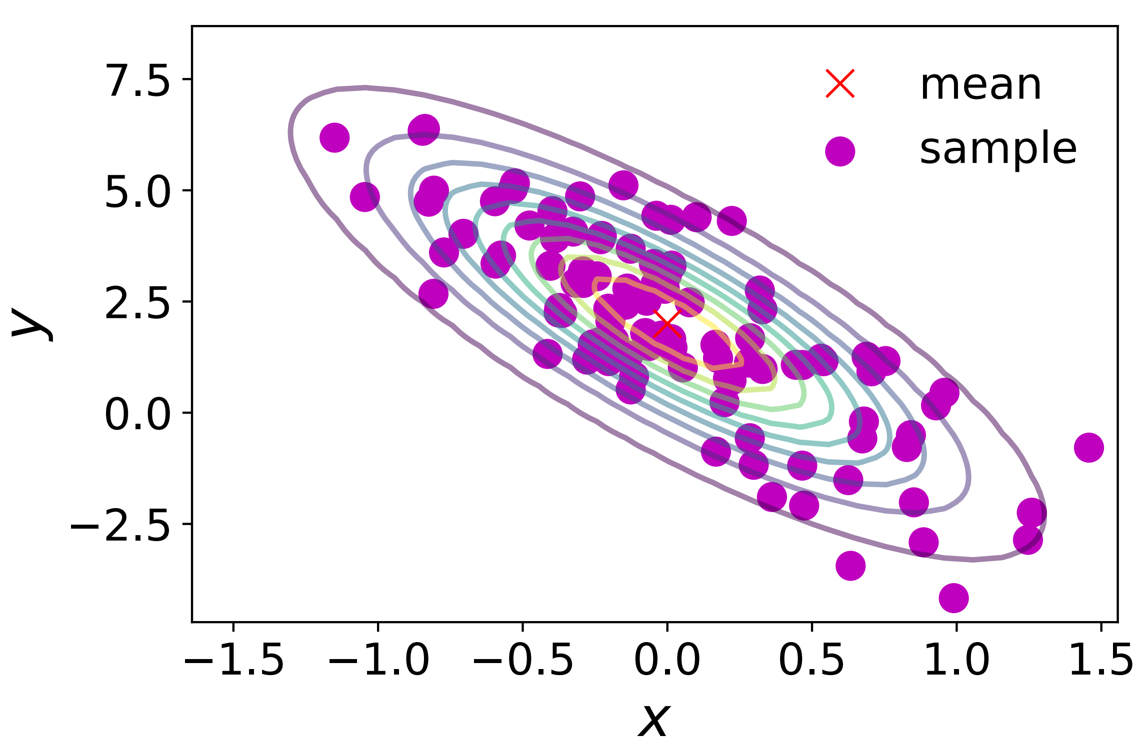
$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

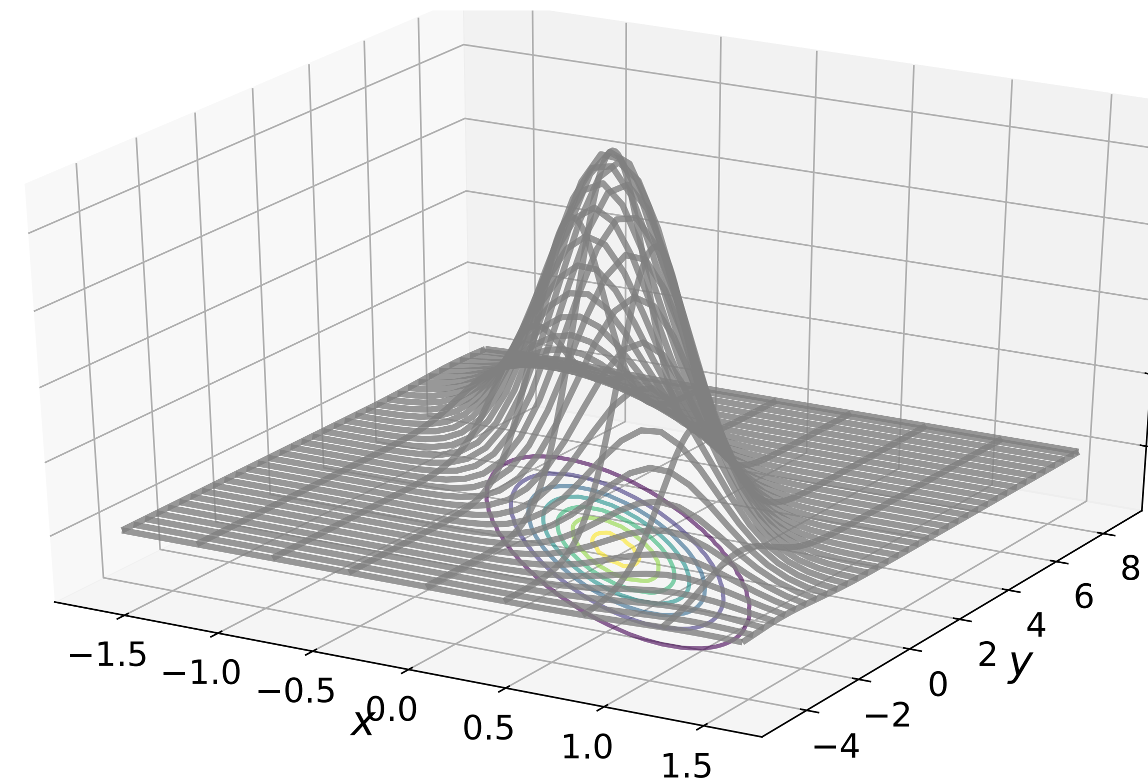


Venn diagrams

Reading: Chapters 1,2 from Bishop "Pattern Recognition and Machine Learning

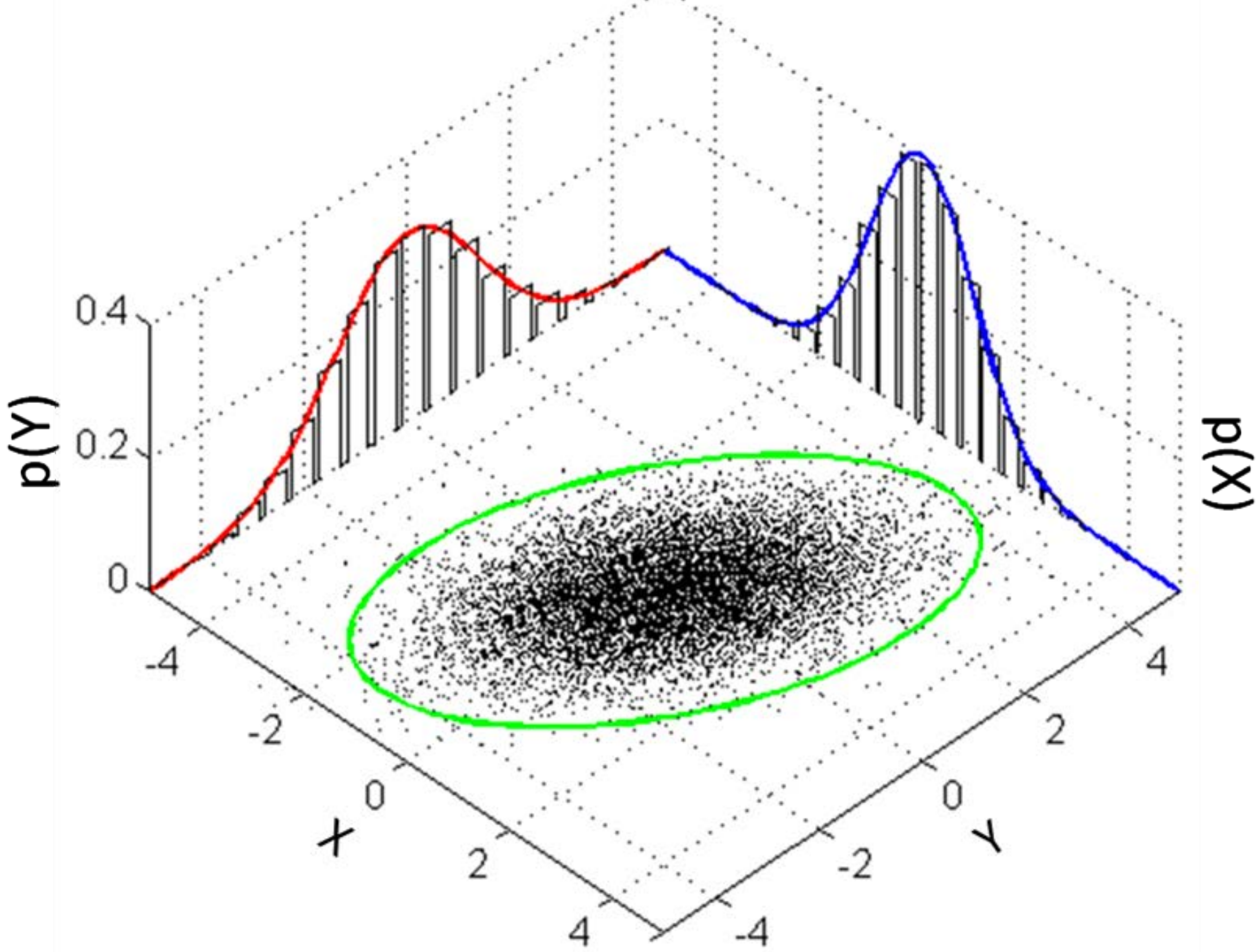
The multivariate Gaussian





$$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)$$

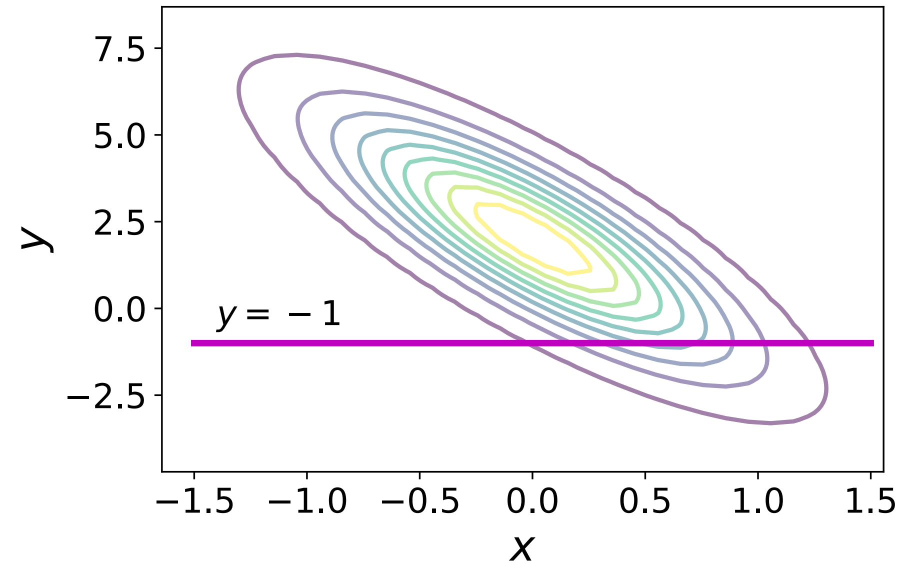
The multivariate Gaussian



Notation	$\mathcal{N}(\boldsymbol{\mu},\,\boldsymbol{\Sigma})$
Parameters	$\boldsymbol{\mu} \in \mathbf{R}^k$ — location $\boldsymbol{\Sigma} \in \mathbf{R}^{k\times k}$ — covariance (positive semi-definite matrix)
Support	$\boldsymbol{x} \in \boldsymbol{\mu} + \mathrm{span}(\boldsymbol{\Sigma}) \subseteq \mathbf{R}^k$
PDF	$\det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$, exists only when $\boldsymbol{\Sigma}$ is positive-definite
Mean	$\boldsymbol{\mu}$
Mode	$\boldsymbol{\mu}$
Variance	$\boldsymbol{\Sigma}$

Marginals and conditionals of a Gaussian

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{bmatrix}\right)$$



Conditional distribution

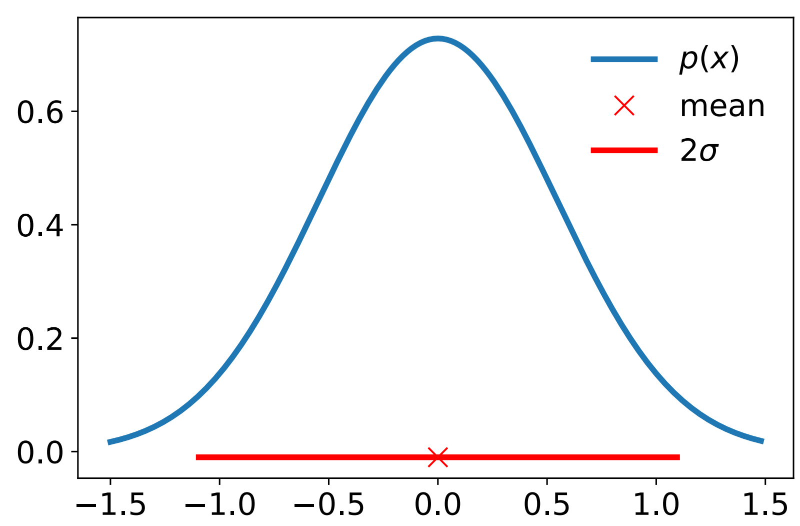
$$p(\boldsymbol{x} \,|\, \boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{\mu}_{x \,|\, y}, \, \boldsymbol{\Sigma}_{x \,|\, y}\right)$$

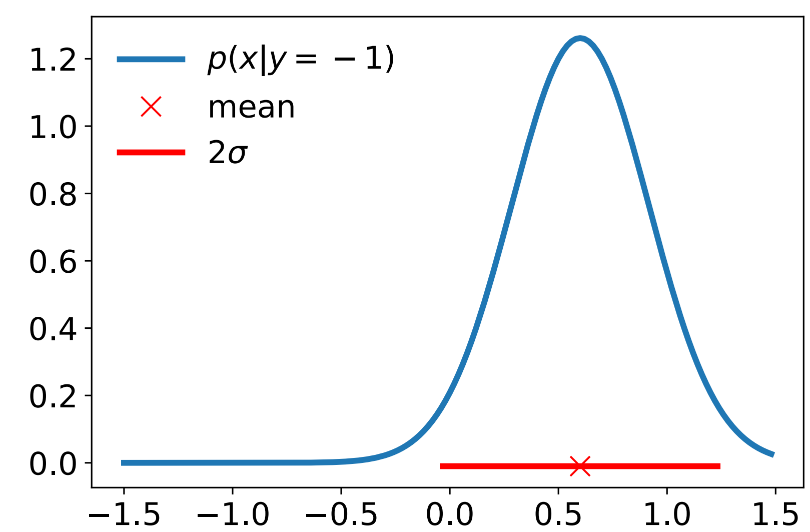
$$\boldsymbol{\mu}_{x \,|\, y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$$

$$\boldsymbol{\Sigma}_{x \,|\, y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} \, .$$

Marginal distribution

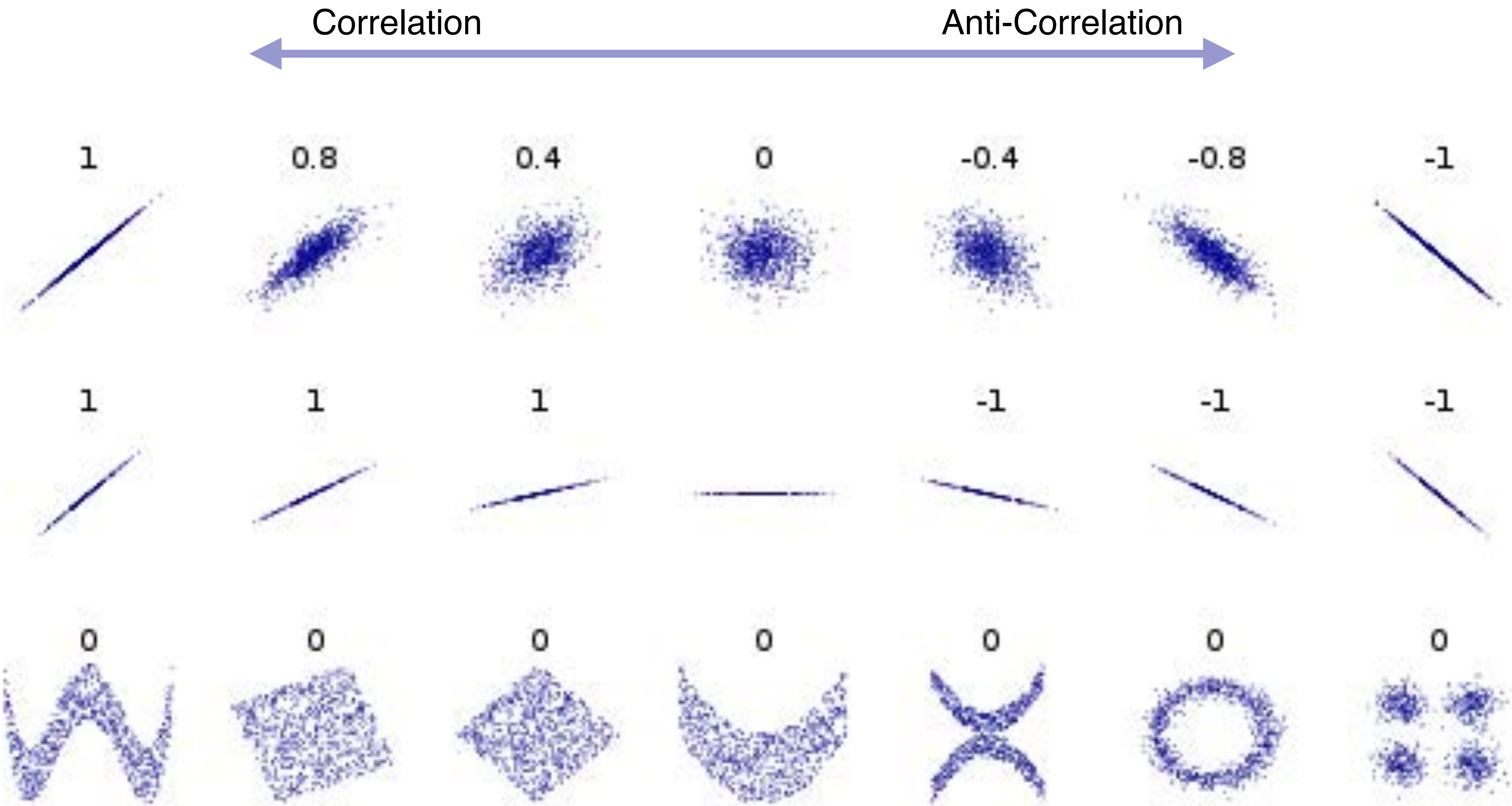
$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} = \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu}_x, \, \boldsymbol{\Sigma}_{xx}\right)$$



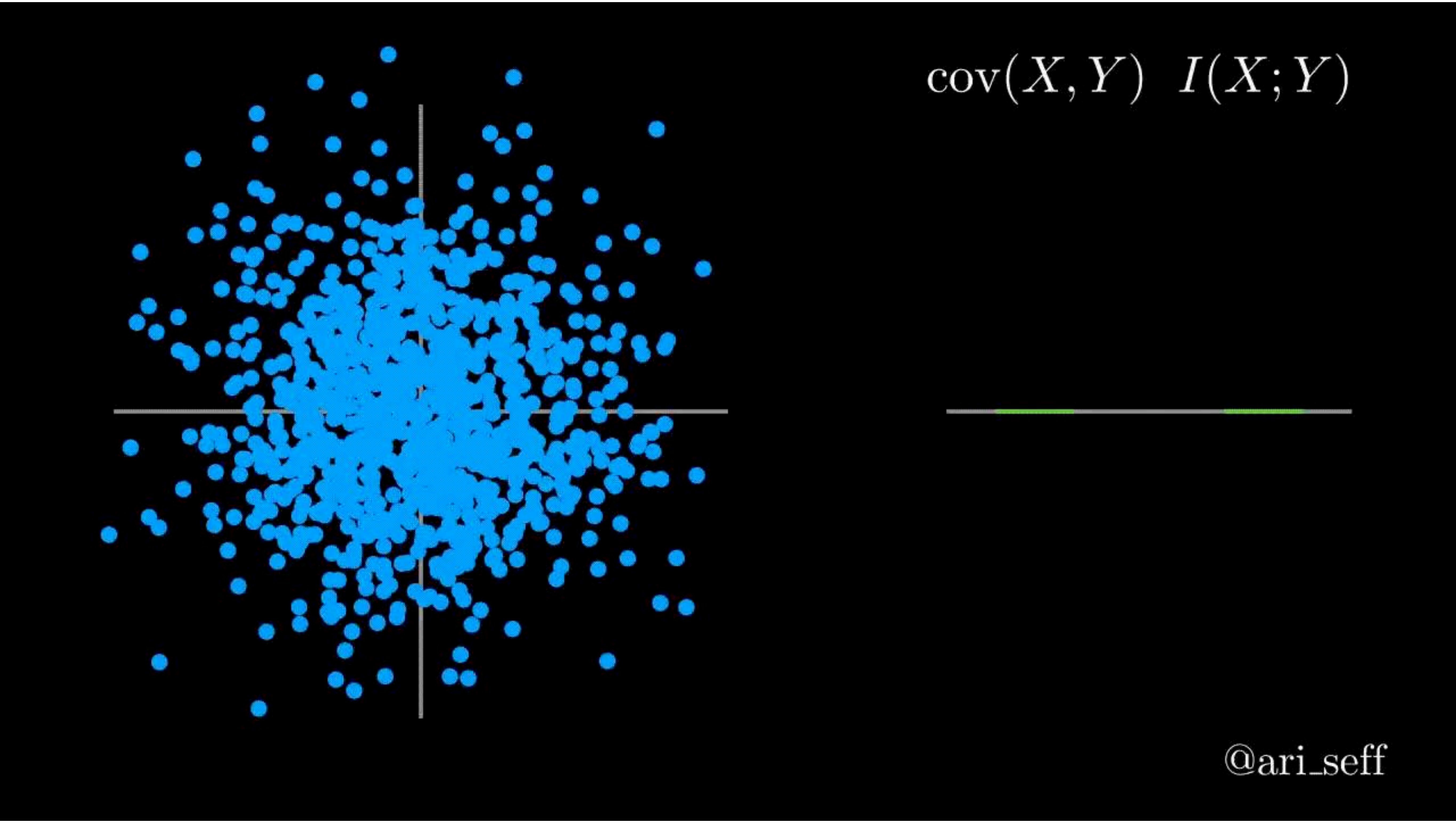


These are unique properties that make the Gaussian distribution very simple and attractive to compute with! It is essentially our main building block for computing under uncertainty.

Correlation and linear dependence

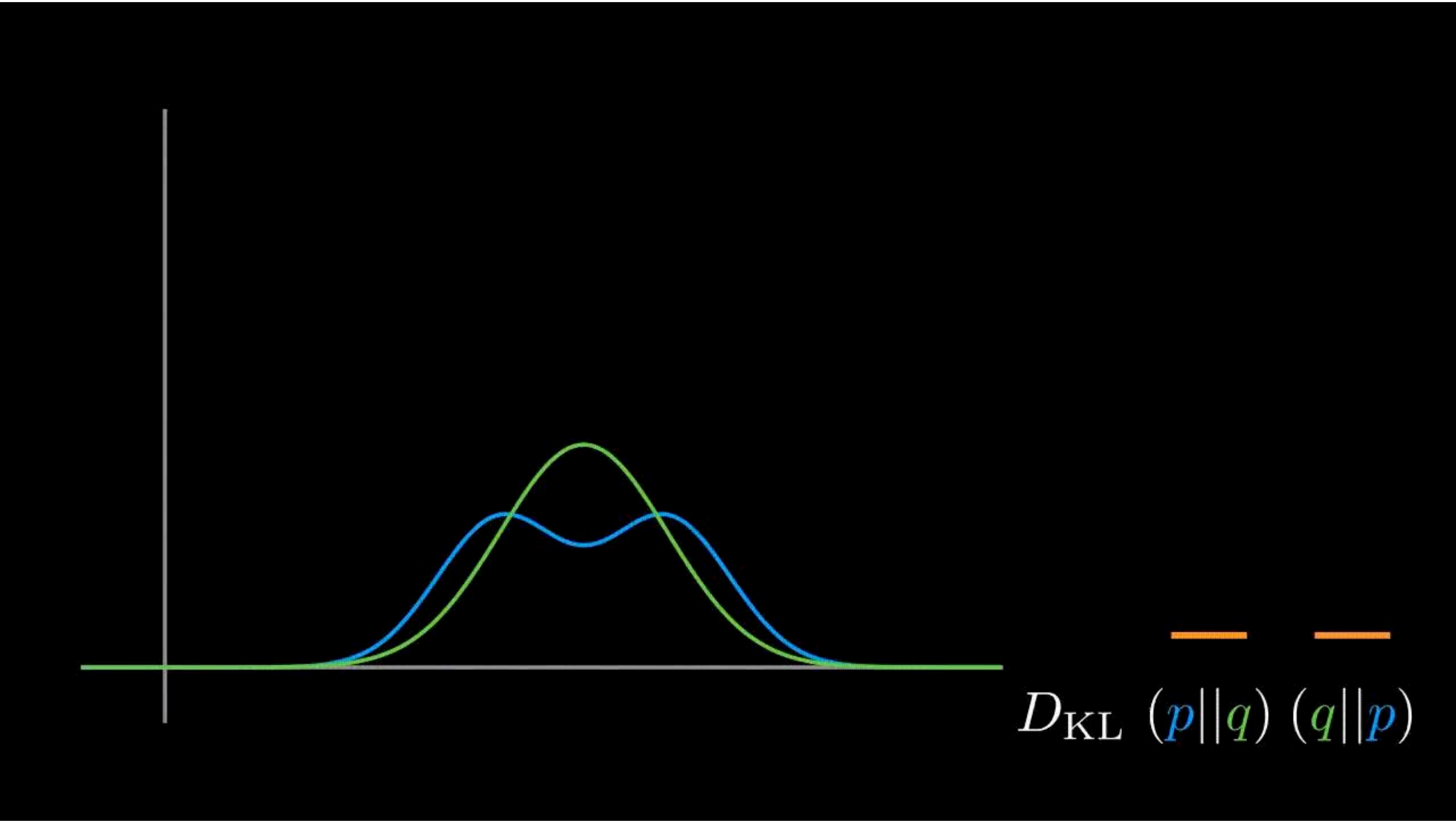


Covariance vs Mutual Information



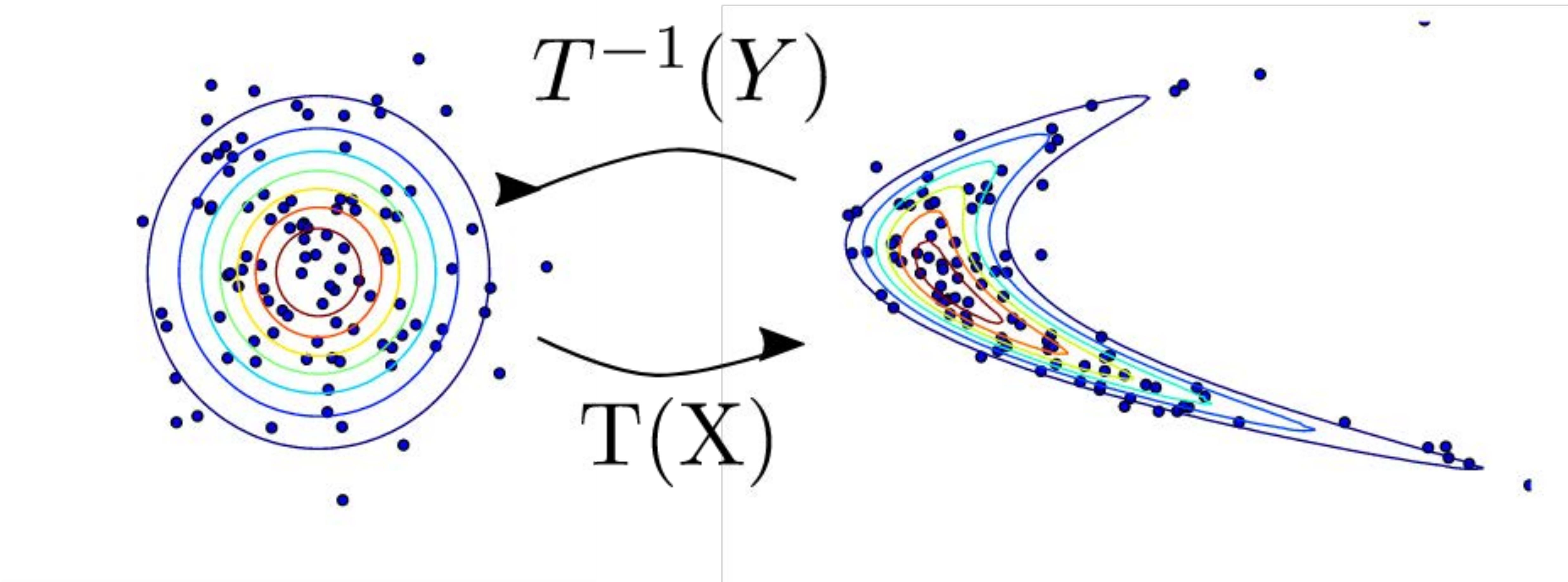
*credit: Ari Seff (Princeton)

Kullbak-Leibler divergence



*credit: Ari Seff (Princeton)

Transformations



Maximum likelihood estimation

$$\theta_{\text{MLE}} = \arg\max_{\theta \in \Theta} p(\mathcal{D}|\theta)$$