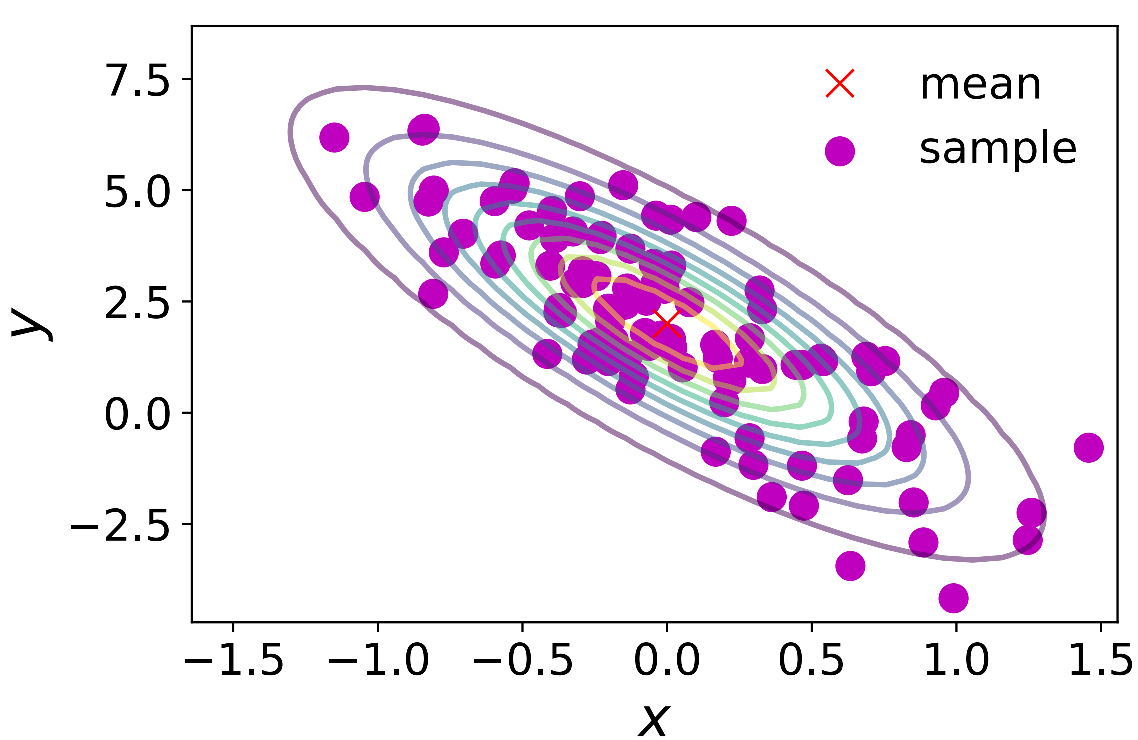
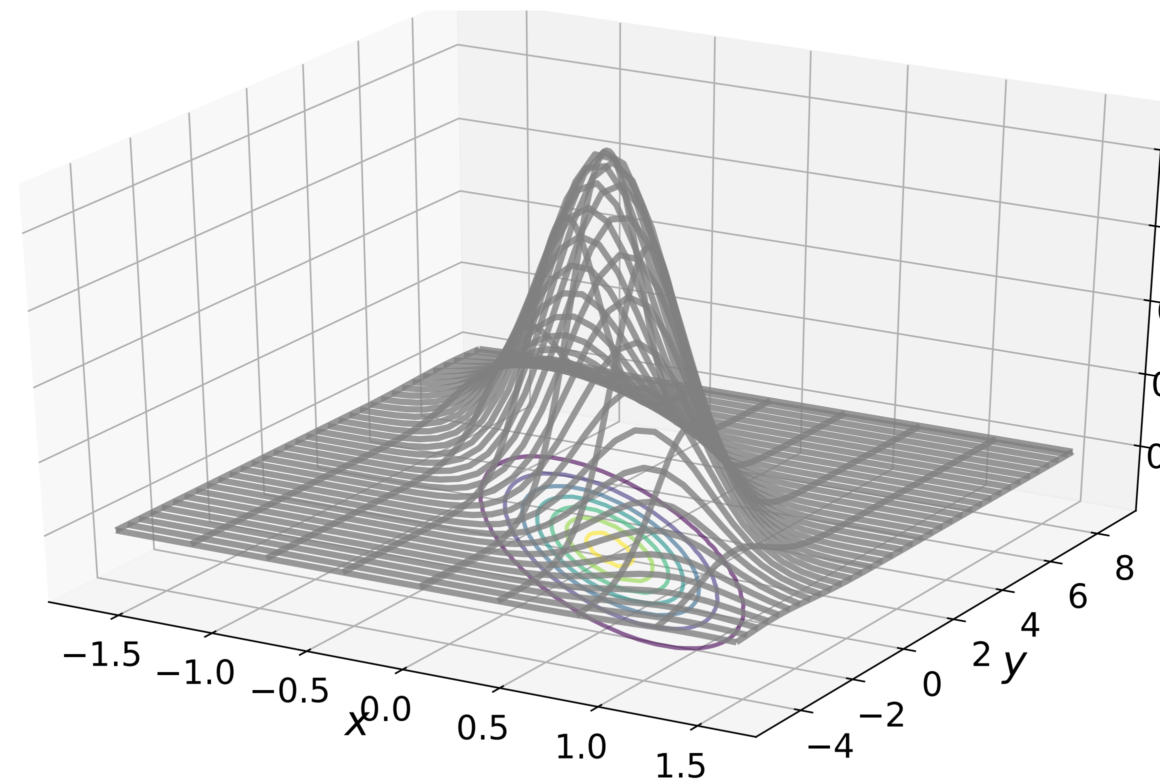
MEAM 4600: AI for Science and Engineering

Lecture #5: Primer on Probability

University *of* Pennsylvania

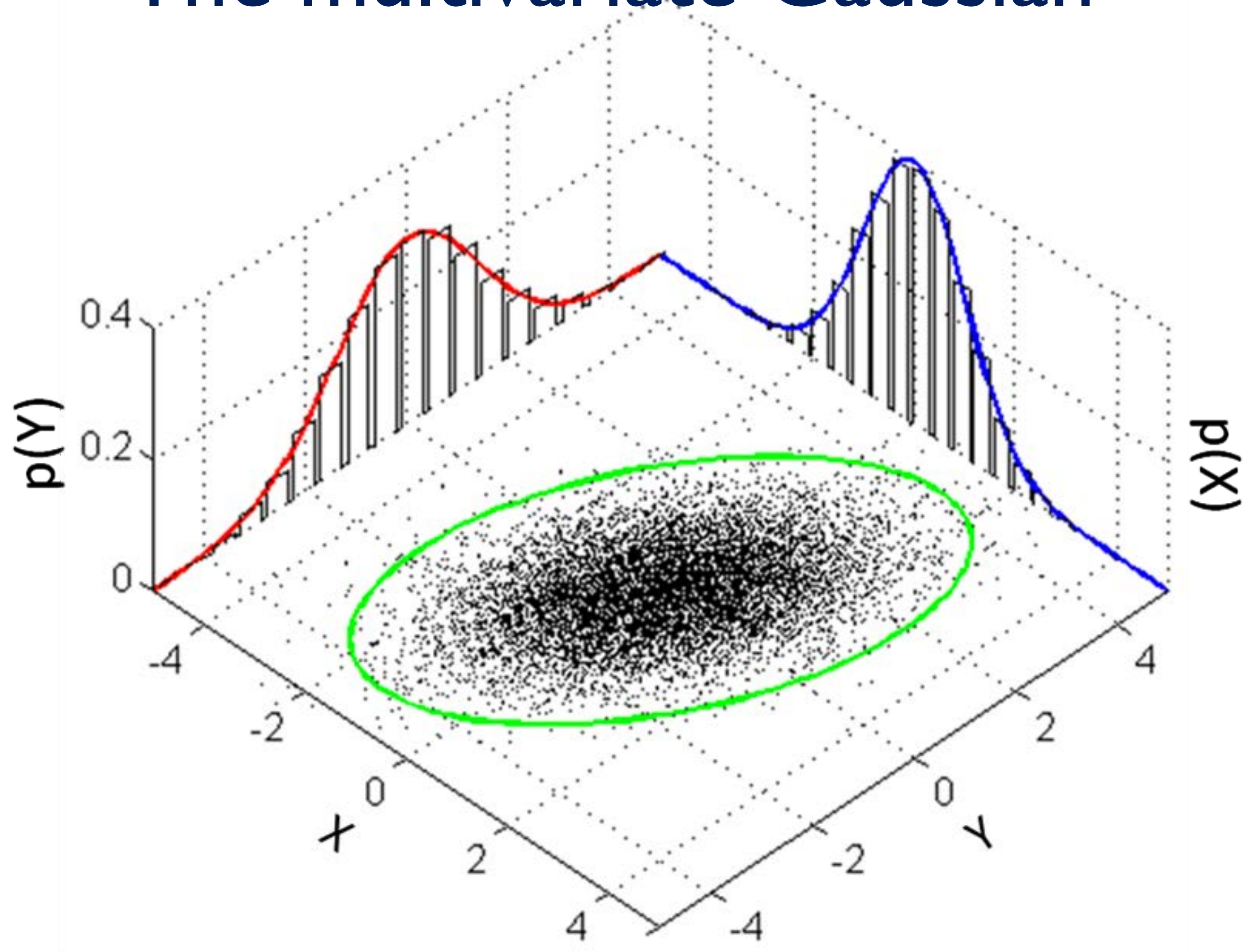
The multivariate Gaussian





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

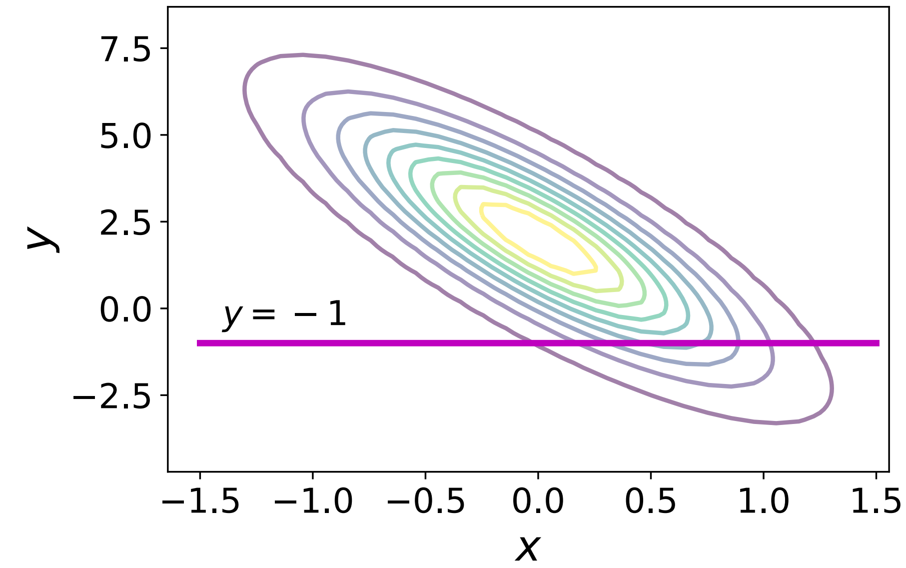
The multivariate Gaussian



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	$\boldsymbol{\Sigma}$

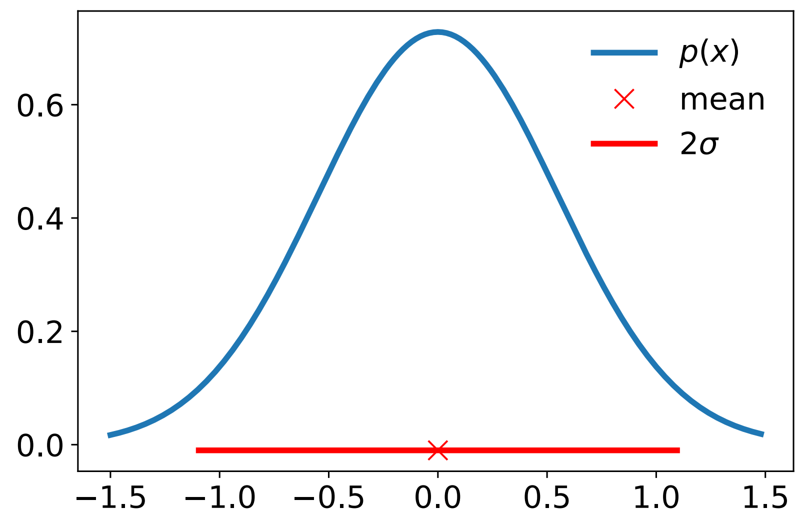
Marginals and conditionals of a Gaussian

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



Marginal distribution

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

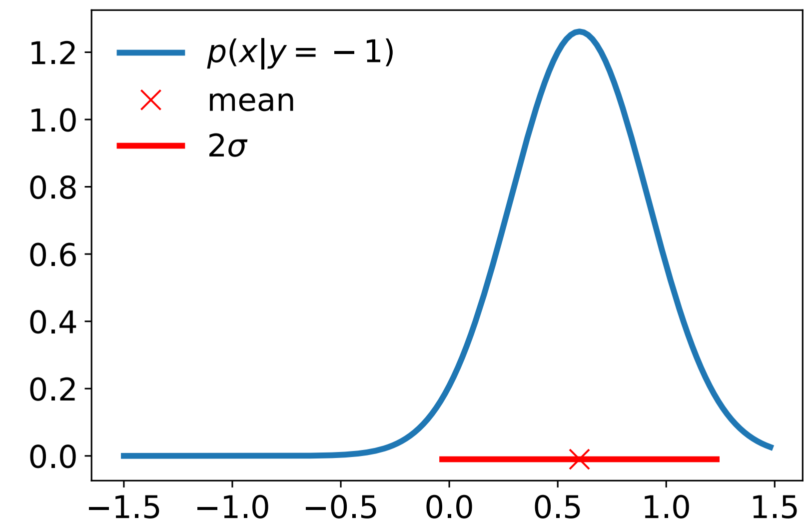


Conditional distribution

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

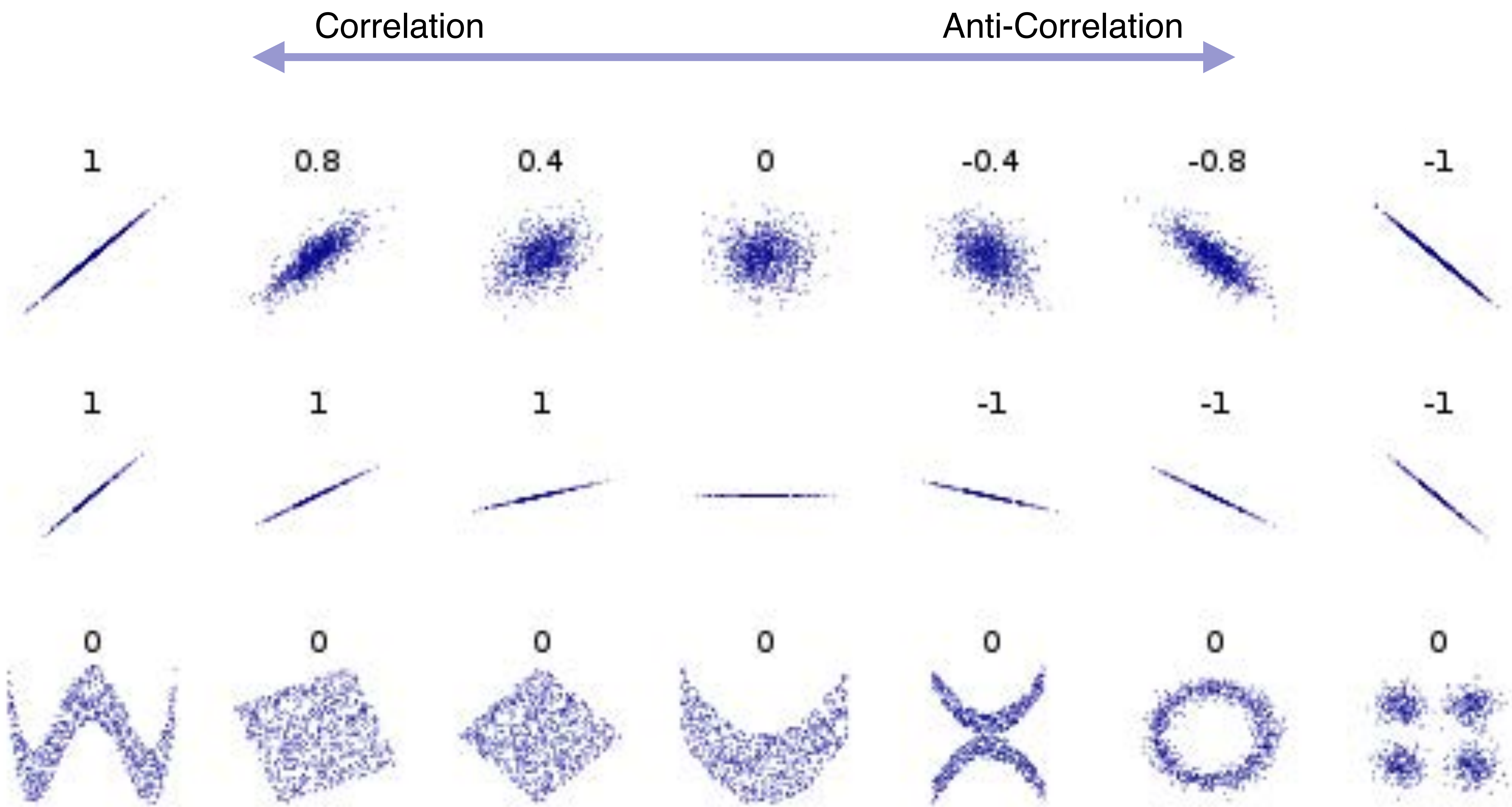
$$\boldsymbol{\mu}_{x\,|\,y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$$

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



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



Entropy and Mutual Information

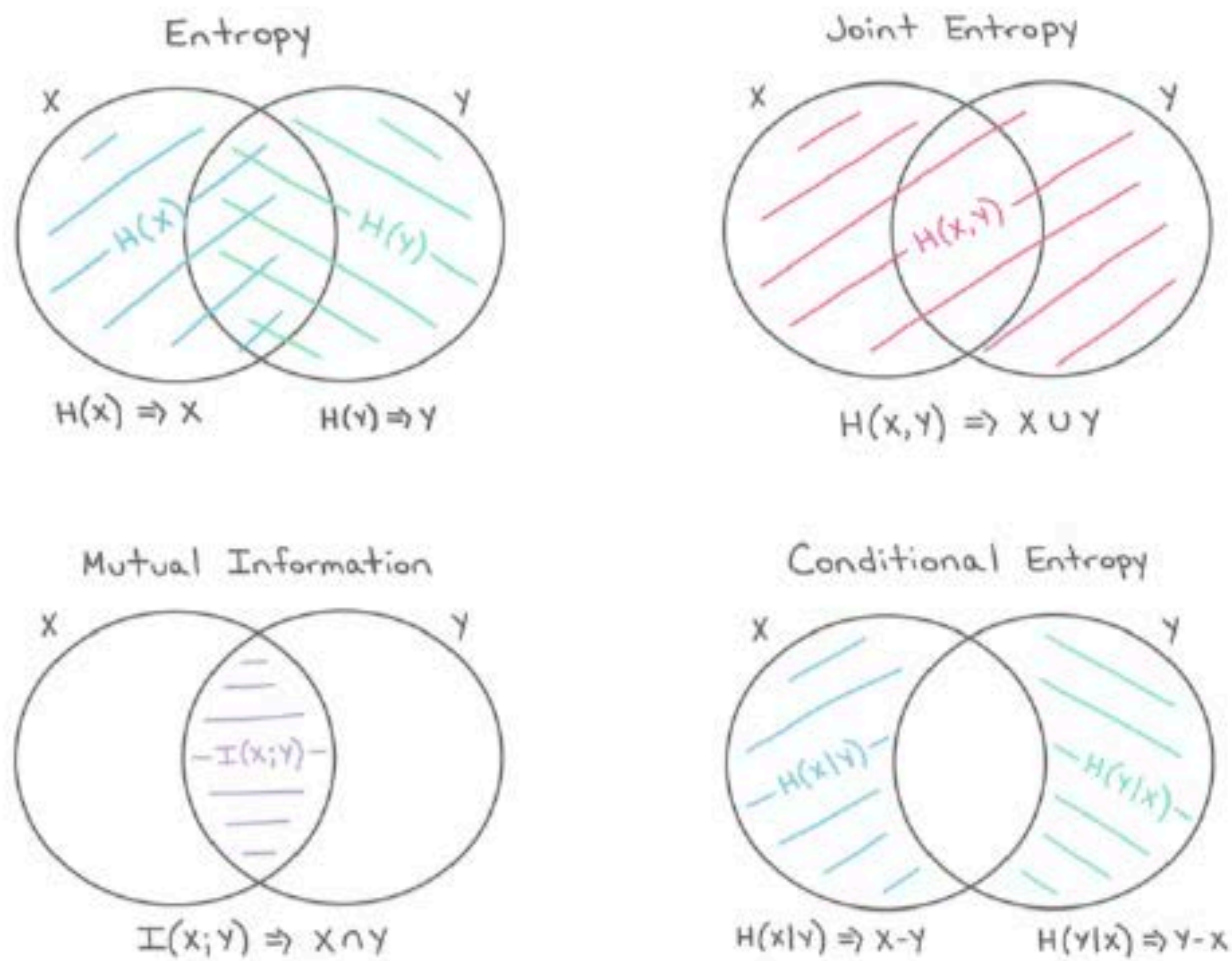


Figure 6.4: The marginal entropy, joint entropy, conditional entropy and mutual information represented as information diagrams. Used with kind permission of Katie Everett.

Forward vs Reverse KL

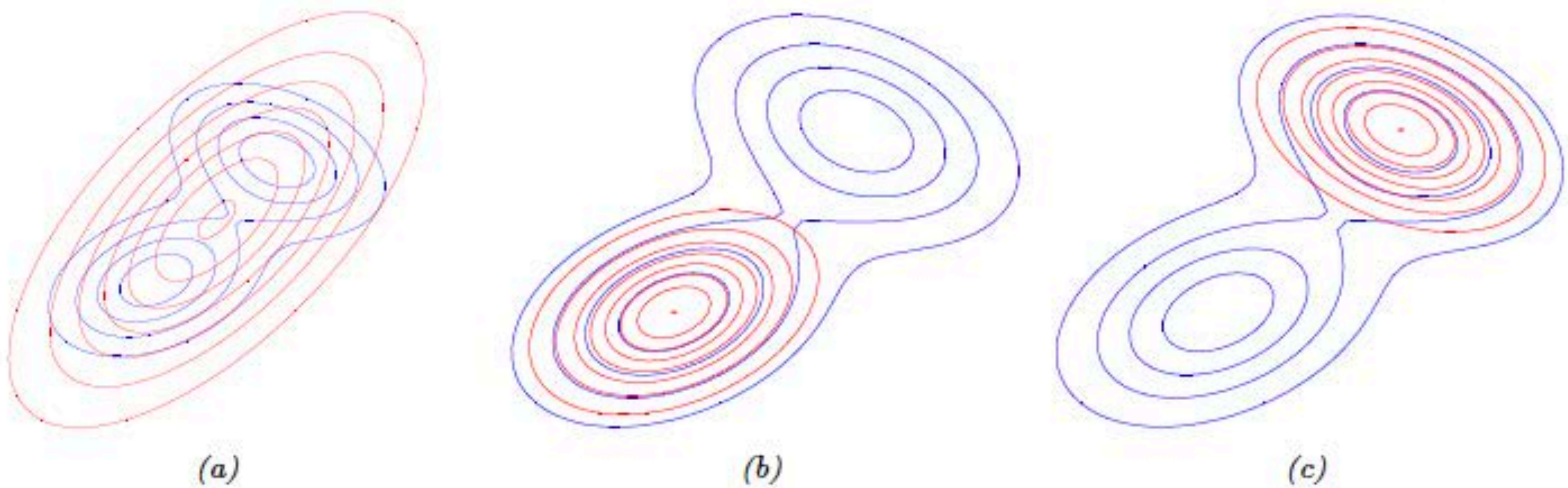


Figure 6.3: Illustrating forwards vs reverse KL on a bimodal distribution. The blue curves are the contours of the true distribution p. The red curves are the contours of the unimodal approximation q. (a) Minimizing forwards KL, $D_{\mathbb{KL}}(p \parallel q)$, wrt q causes q to "cover" p. (b-c) Minimizing reverse KL, $D_{\mathbb{KL}}(q \parallel p)$ wrt q causes q to "lock onto" one of the two modes of p. Adapted from Figure 10.3 of [Bis06]. Generated by KLfwdReverseMixGauss.ipynb.

Covariance vs Mutual Information