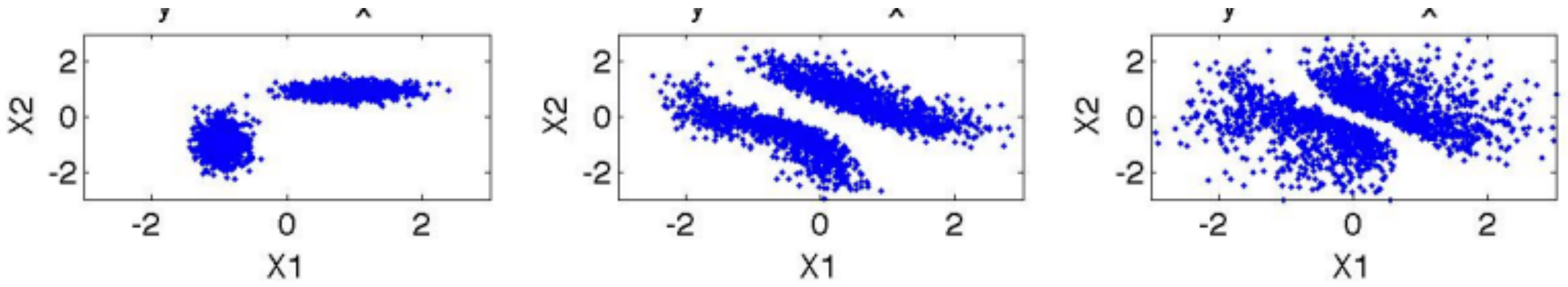
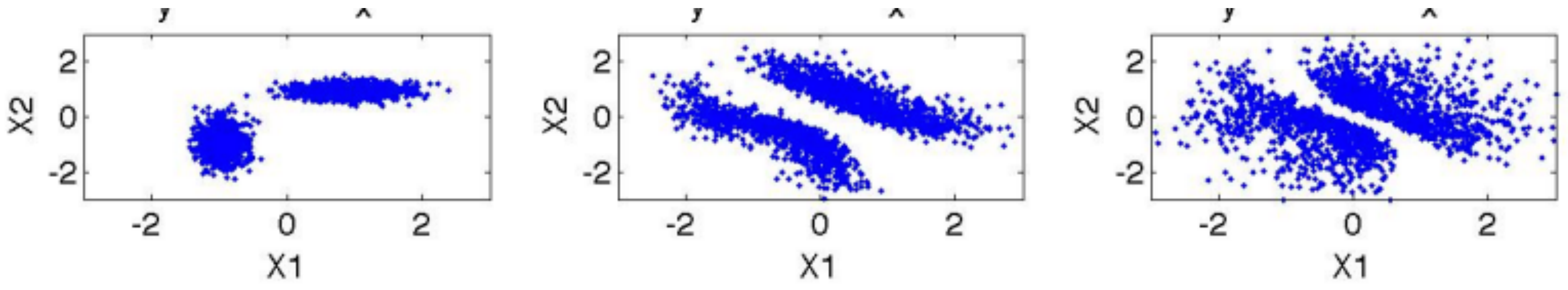
Normalizing Flow history



2010
Tabak, Esteban G., and Cristina V. Turner.
"A family of nonparametric density estimation algorithms."

Normalizing Flow history



2010
Tabak, Esteban G., and Cristina V. Turner.
"A family of nonparametric density estimation algorithms."

3.1 Regularizing the Transformations

We will train the model on data by minimizing an objective composed of several parts:

Divergence Penalty $\mathcal{D}(\boldsymbol{\Psi})$: This determines the fit of the current encoding transformation. It forces the marginal densities of the empirical distribution of the representation-space data to match a target distribution of our choice, by penalizing divergence from it.

Invertibility Measure $\mathcal{I}(\boldsymbol{\Theta})$: This ensures the invertibility of $\boldsymbol{f}_{\boldsymbol{\Theta}}(\cdot)$ by penalizing poorly-conditioned transformations.

Reconstruction Loss $\mathcal{R}(\boldsymbol{\Theta}, \boldsymbol{\Psi})$: This jointly penalizes the encoder $\boldsymbol{g}_{\boldsymbol{\Psi}}(\cdot)$ and decoder $\boldsymbol{f}_{\boldsymbol{\Theta}}(\cdot)$ to ensure that $\boldsymbol{g}_{\boldsymbol{\Psi}}(\boldsymbol{y}) \approx \boldsymbol{f}_{\boldsymbol{\Theta}}^{-1}(\boldsymbol{y})$ on the data.

Each of these participates in the overall objective given by:

$$C(\boldsymbol{\Theta}, \boldsymbol{\Psi}) = \mu_{\mathcal{D}}\mathcal{D}(\boldsymbol{\Theta}) + \mu_{\mathcal{I}}\mathcal{I}(\boldsymbol{\Psi}) + \mu_{\mathcal{R}}\mathcal{R}(\boldsymbol{\Theta}, \boldsymbol{\Psi}) \; , \tag{6}$$

where $\mu_{\mathcal{I}}, \mu_{\mathcal{D}}, \mu_{\mathcal{R}} \in \mathbb{R}$ are the weights of each term. We will examine each of these terms in more detail in the proceeding sections.

2013
Rippel, Oren, and Ryan Prescott Adams.
“High-dimensional probability estimation with deep density models."