

Siddhant Mishra-Sharma (MIT/AI FI) Summer School

162

2

7

Rate-distortion theory: *neutral compression*

Autoencoding is a form of (neural) compression!

$$\text{ELBO} = \left\langle \log p_{\vartheta}(x \mid z) \right\rangle_{q_{\varphi}} - D_{\text{KL}}\left(q_{\varphi}(z \mid x) \parallel p(z)\right)$$



Rate

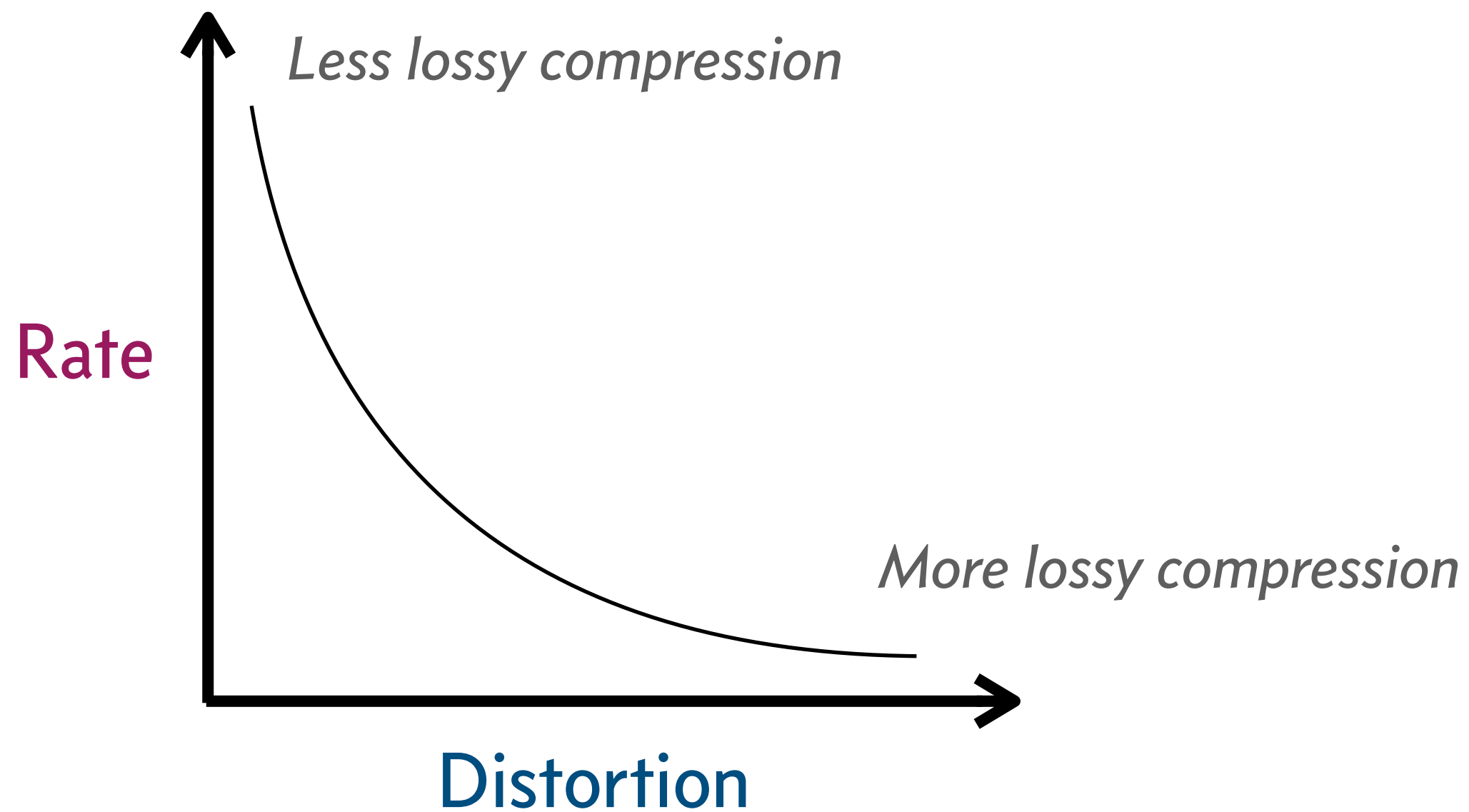
“Amount of compression”



Distortion

“Information loss”

*Rate-distortion curve
quantified this tradeoff*



Rate-distortion theory: *neural compression*

Autoencoding is a form of (neural) compression!

$$\text{ELBO} = \underbrace{\langle \log p_{\vartheta}(x | z) \rangle_{q_{\varphi}}}_{\text{Distortion}} - \underbrace{D_{\text{KL}}(q_{\varphi}(z | x) \| p(z))}_{\text{Rate}}$$

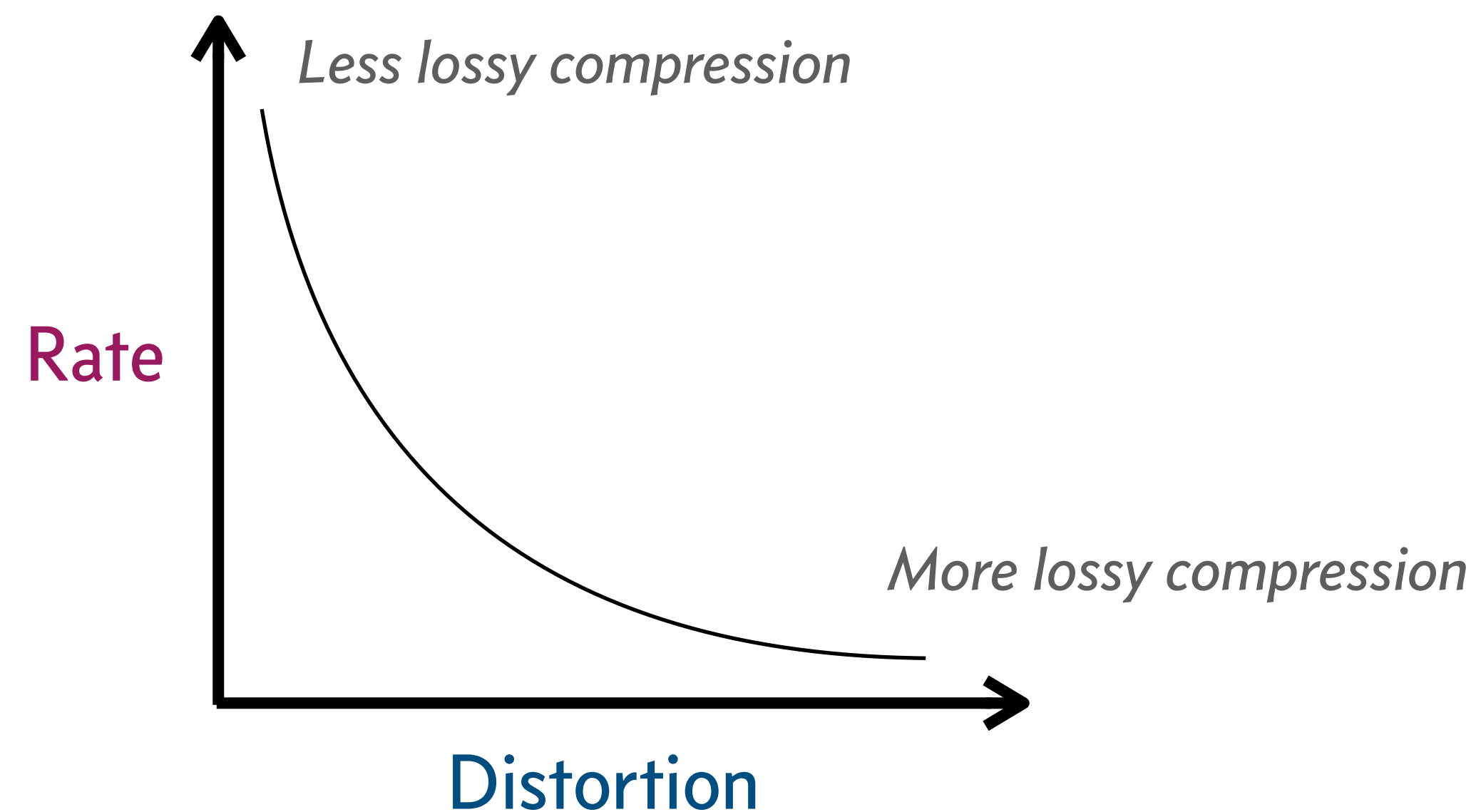
Distortion

“Information loss”

Rate

“Amount of compression”

*Rate-distortion curve
quantified this tradeoff*



Controlling compression and disentanglement: β -VAEs

$$\text{ELBO} = \underbrace{\left\langle \log p_{\vartheta}(x \mid z) \right\rangle_{q_{\varphi}}}_{\text{Distortion}} - \underbrace{\beta \cdot D_{\text{KL}} \left(q_{\varphi}(z \mid x) \parallel p(z) \right)}_{\text{Rate}}$$

Distortion

Rate