

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

170

29

Neural compression: *Rate-distortion theory*

Autoencoding is a form of (neural) compression!

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



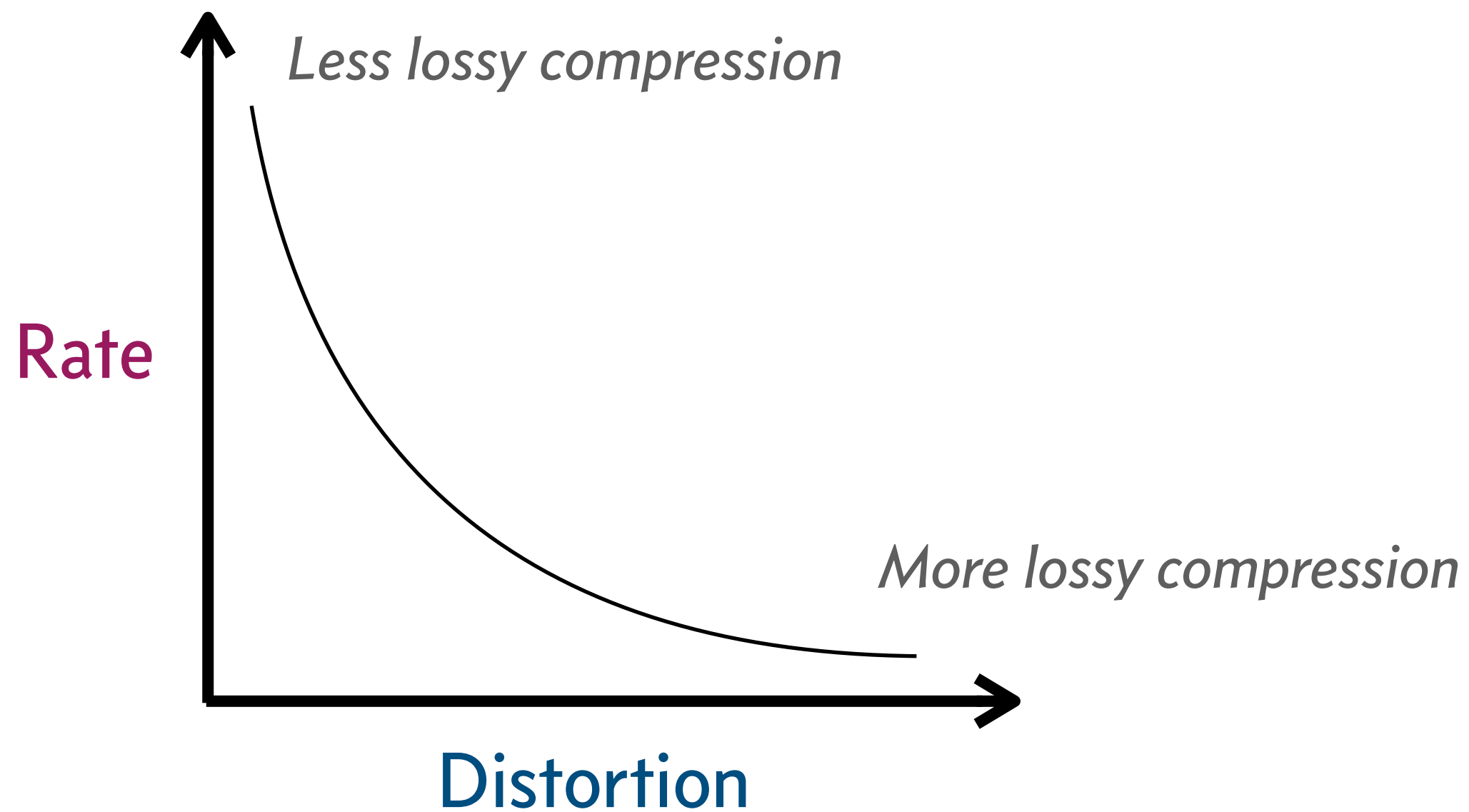
Rate

“Amount of compression”

Distortion

“Reconstruction loss”

*Rate-distortion curve
quantified this tradeoff*



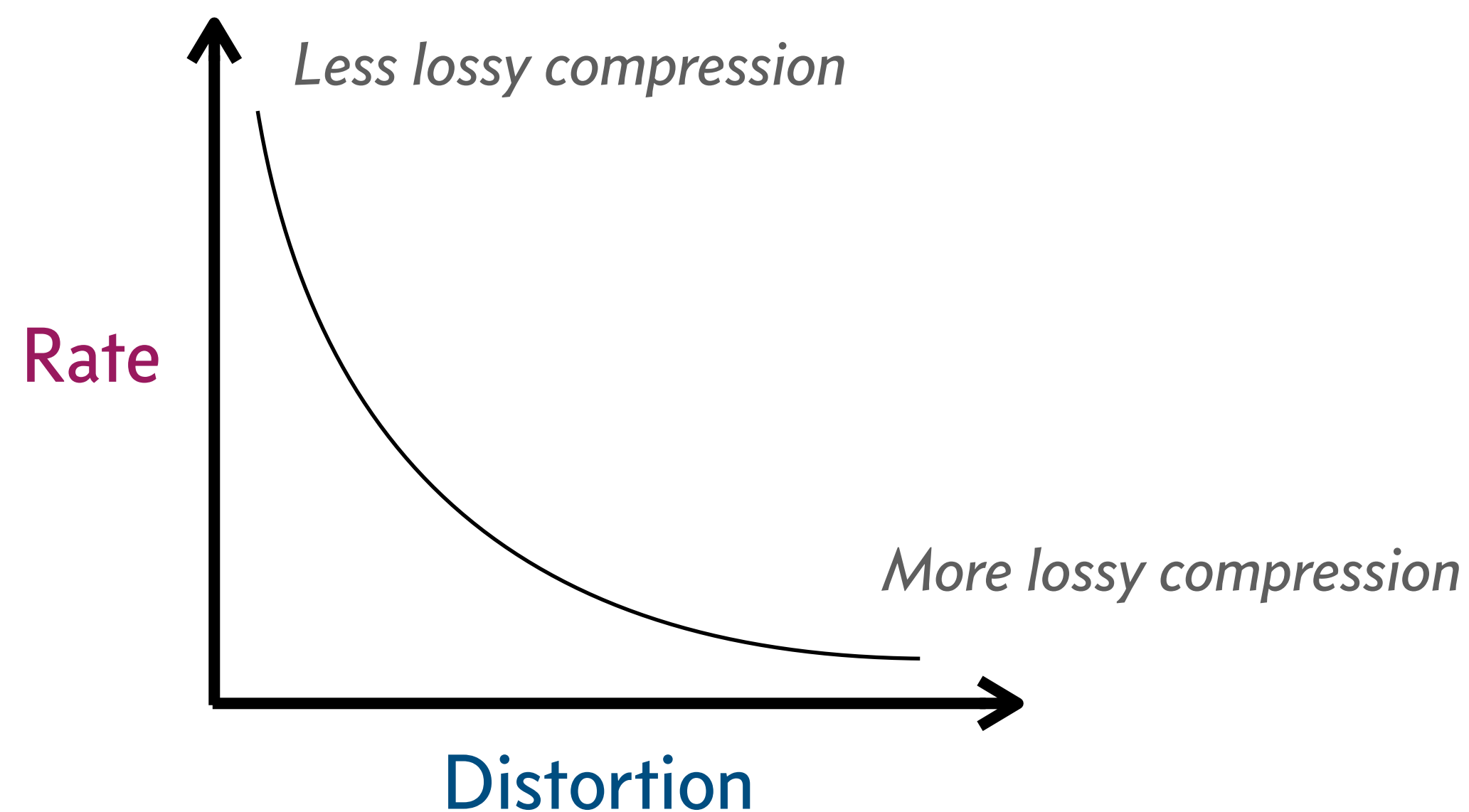
Neural compression: *Rate-distortion theory*

Autoencoding is a form of (neural) compression!

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

“Reconstruction loss” “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