/66

15


I AM GOING TO
TRAIN A GENERATIVE MODEL
BY MAXIMIZING
THE LIKELIHOOD OF
TRAINING DATA, RIGHT?

Latent-variable modeling

Maximum-likelihood training?

$$\begin{aligned}\vartheta^* &= \arg\max_{\vartheta} p_{\vartheta}(x) \\ &= \arg\max_{\vartheta} \int p_{\vartheta}(x \mid z) p(z)\, \mathrm{d}z \\ &= \arg\max_{\vartheta} \left\langle p_{\vartheta}(x \mid z) \right\rangle_{p(z)}\end{aligned}$$

Difficult to build a good estimator!

Latent-variable modeling

Maximum-likelihood training?

$$\vartheta^* = \arg\max_{\vartheta} p_\vartheta(x)$$
$$= \arg\max_{\vartheta} \int p_\vartheta(x \mid z) p(z)\, \mathrm{d}z$$
$$= \arg\max_{\vartheta} \left\langle p_\vartheta(x \mid z) \right\rangle_{p(z)}$$

Difficult to build a good estimator!

KL-divergence

A measure of similarity between two probability distributions

$$D_{\mathrm{KL}}(Q\|P) = \int_{-\infty}^{\infty} \mathrm{d}x\, q(x)\, \log\left(\frac{q(x)}{p(x)}\right)$$

$$= \left\langle \log \frac{q(x)}{p(x)} \right\rangle_{x\sim q(x)}$$

Formally: *expected excess "surprise" from using P as a model when the actual distribution is Q*