/70

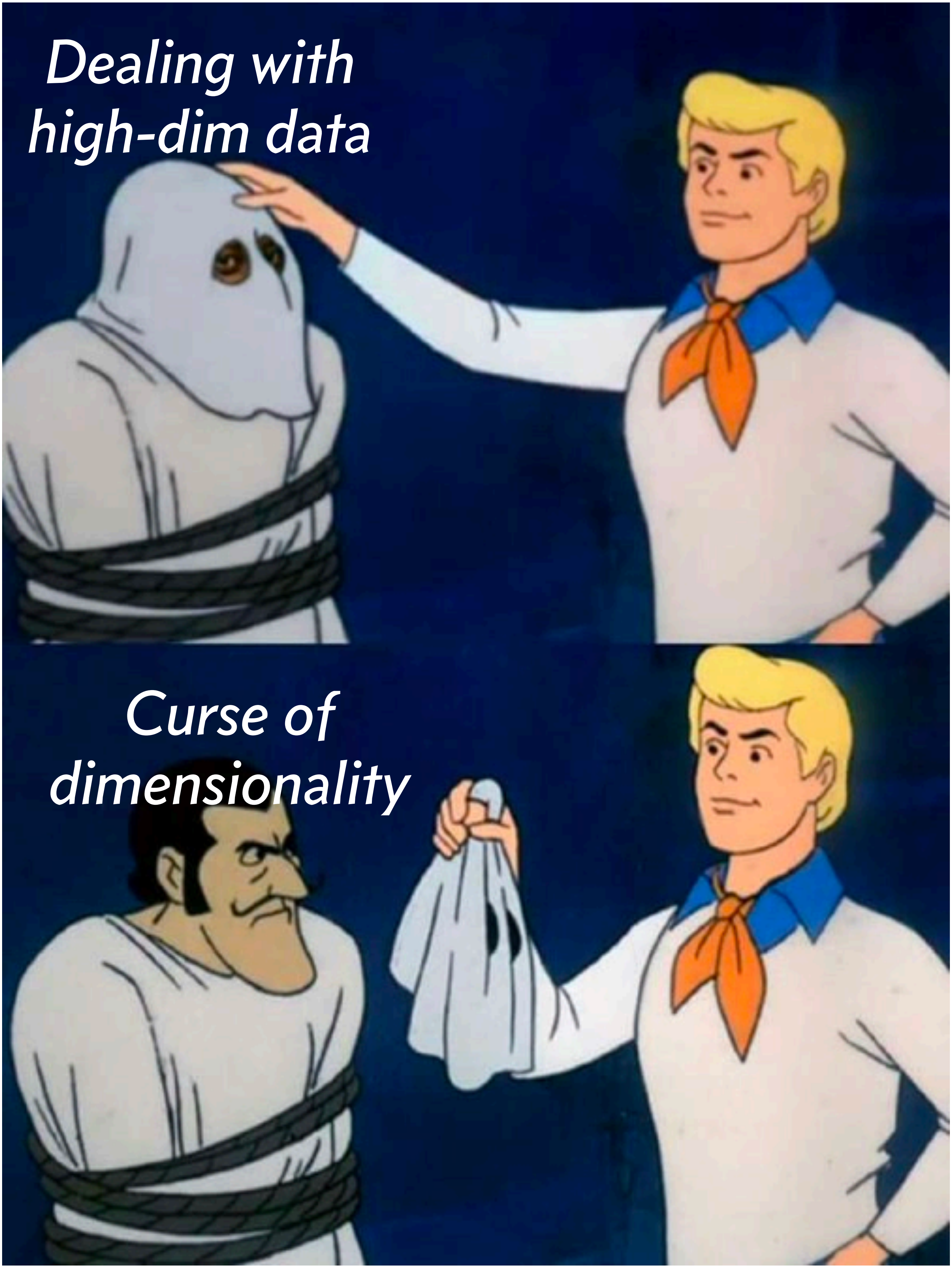
18

Latent-variable modeling

Maximum-likelihood training?

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

Difficult to build a good estimator!


Dealing with high-dim data
Curse of dimensionality

The intractability of $p(x)$ is closely related to the intractability of the *posterior* $p(z \mid x)$

$$\boxed{p(z \mid x)} = \frac{\boxed{p(x, z)}}{\boxed{p(x)}} \qquad \text{(Bayes' theorem)}$$

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!

The intractability of $p(x)$ is closely related to the intractability of the *posterior* $p(z \mid x)$

$$p(z \mid x) = \frac{p(x, z)}{p(x)} \quad \text{(Bayes' theorem)}$$

Kullback–Leibler (KL) divergence

A measure of similarity between two probability distributions

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

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