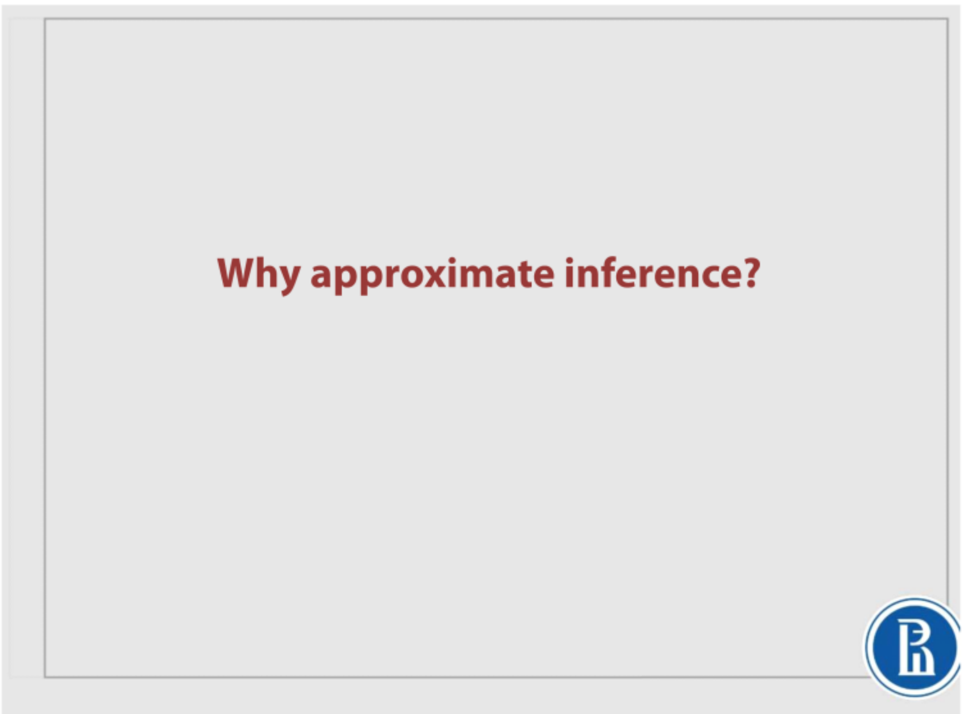
Variational Inference

Monday, June 25, 2018 11:47 PM

why approximate



Analytical inference

Why compute Approx.?

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)}$$

Analytical inference

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)}$$

- Easy for conjugate priors

Analytical inference

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)}$$

- Easy for conjugate priors
    - Hard otherwise

Analytical inference

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)}$$

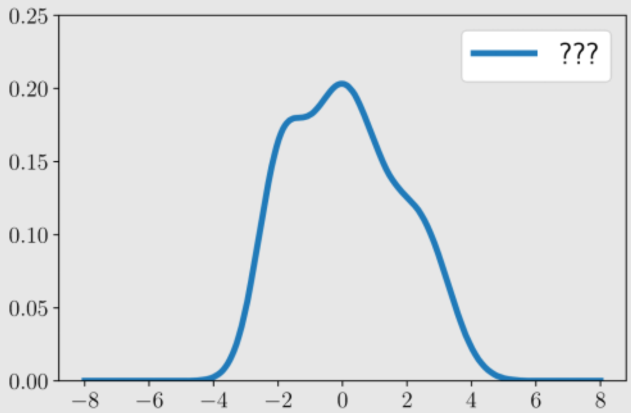
- Easy for conjugate priors
- Hard otherwise

variational.
auto-encoder

Example: $p(x|z) = \mathcal{N}(x|\mu(z), \sigma^2(z))$

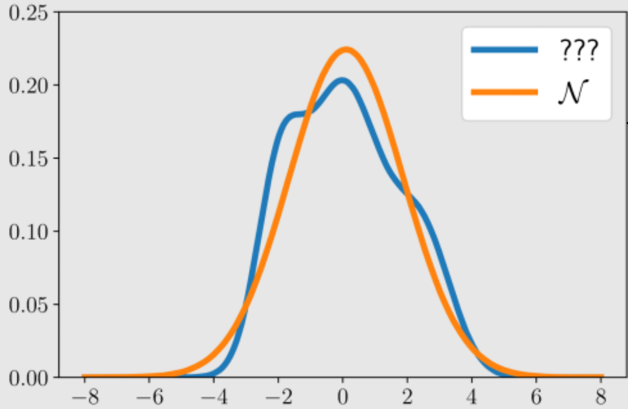
Neural networks

Do we need exact posterior?



Not belong to known distributions.

Do we need exact posterior?



→ Approximate with Gaussian.

$\boxed{\text{Task}}$ Find the best approx.

Variational inference

1. Select a family of distributions Q

   Example: $\mathcal{N}\left(\mu, \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \\ & & & \sigma_d^2 \end{pmatrix}\right)$

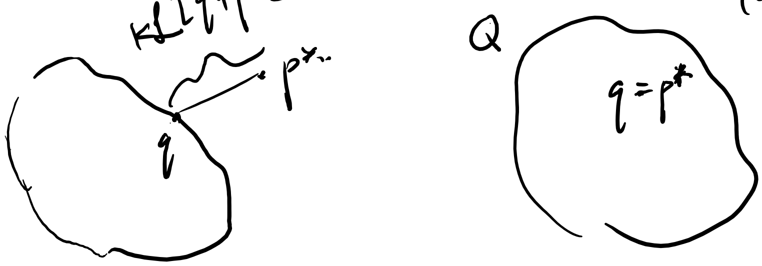
2. Find best approximation $q(\underline{z})$ of $p^*(z)$

$$KL\left[q(z) \,\|\, p^*(z)\right] \to \min_{q \in Q}$$

Choice of variational family.

$KL[q \| p^*]$.

larger Q. p^* in Q.



$p^* \notin Q$. $p^* \in Q$.

Smaller set of distribution Q $\Rightarrow$ Larger approx ~~of~~ error

Unnormalized distribution one problem need to compute p^*.

$$p^*(z) = p(z \mid x) = \frac{p(x \mid z)\, p(z)}{p(x)} = \frac{\hat{p}(z)}{Z} \qquad \boxed{\text{optimization}}$$

$$KL\left[q(z) \,\Big\|\, \frac{\hat{p}(z)}{Z}\right] = \int q(z) \log \frac{q(z)}{\hat{p}(z)/Z}\, dz$$

???

$$= \int q(z) \log \frac{q(z)}{\hat{p}(z)} + \int q(z) \log Z\, dz$$

$$KL\left[q(z) \,\|\, \hat{p}(z)\right] \to \min_{z}$$

varinf2

Variational Inference

Variational inference

1. Select a family of distributions Q

Variational inference

1. Select a family of distributions Q

Example: $\mathcal{N}(\mu, \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & 0 \\ 0 & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix})$

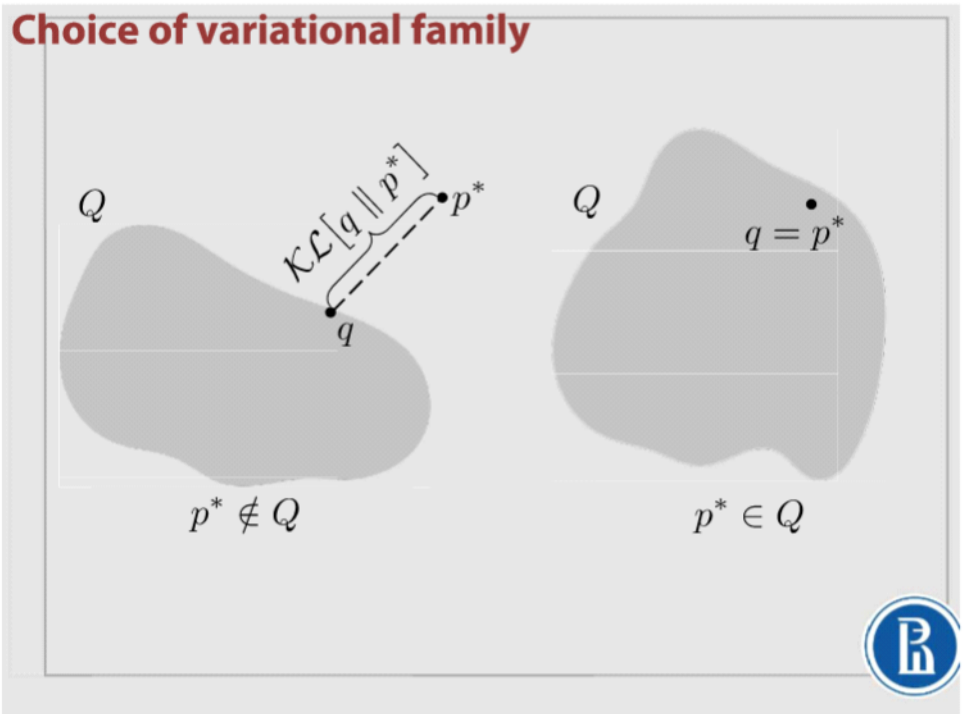
Variational inference

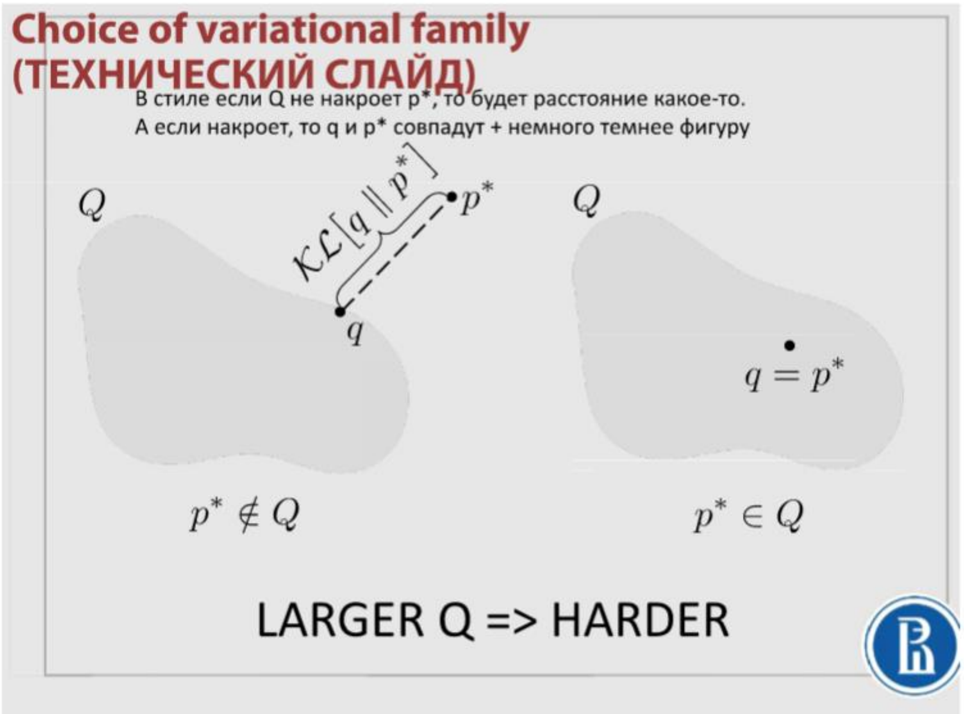
1. Select a family of distributions Q

Example: $\mathcal{N}(\mu, \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & 0 \\ 0 & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix})$

2. Find best approximation $q(z)$ of $p^*(z)$:

$$\mathcal{KL}\left[q(z) \parallel p^*(z)\right] \to \min_{q \in Q}$$


Choice of variational family
Q
$KL[q \,\|\, p^*]$
p^*
q
$p^* \notin Q$
Q
$q = p^*$
$p^* \in Q$


Choice of variational family
(ТЕХНИЧЕСКИЙ СЛАЙД)
В стиле если Q не накроет p*, то будет расстояние какое-то.
А если накроет, то q и p* совпадут + немного темнее фигуру
Q
$\mathcal{KL}[q \,\|\, p^*]$
p^*
q
Q
$q = p^*$
$p^* \notin Q$
$p^* \in Q$
LARGER Q => HARDER

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

$$\mathcal{KL}\left[q(z) \parallel \frac{\widehat{p}(z)}{Z}\right] =$$

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

$$\mathcal{KL}\big[q(z) \parallel \frac{\widehat{p}(z)}{Z}\big] = \int q(z) \log \frac{q(z)}{\widehat{p}(z)/Z} dz$$

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

$$\mathcal{KL}\left[q(z) \parallel \frac{\widehat{p}(z)}{Z}\right] = \int q(z) \log \frac{q(z)}{\widehat{p}(z)/Z} dz$$

$$= \int q(z) \log \frac{q(z)}{\widehat{p}(z)} dz + \int q(z) \log Z dz$$

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

$$\begin{aligned}\mathcal{KL}\big[q(z) \parallel \frac{\widehat{p}(z)}{Z}\big] &= \int q(z) \log \frac{q(z)}{\widehat{p}(z)/Z} dz \\ &= \int q(z) \log \frac{q(z)}{\widehat{p}(z)} dz + \int q(z) \log Z dz \\ &= \mathcal{KL}\big[q(z) \parallel \widehat{p}(z)\big] + \log Z\end{aligned}$$

Unnormalized distribution

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$

Optimization:

$$\mathcal{KL}\left[q(z) \parallel \frac{\widehat{p}(z)}{Z}\right] = \int q(z) \log \frac{q(z)}{\widehat{p}(z)/Z} dz$$

$$= \int q(z) \log \frac{q(z)}{\widehat{p}(z)} dz + \int q(z) \log Z dz$$

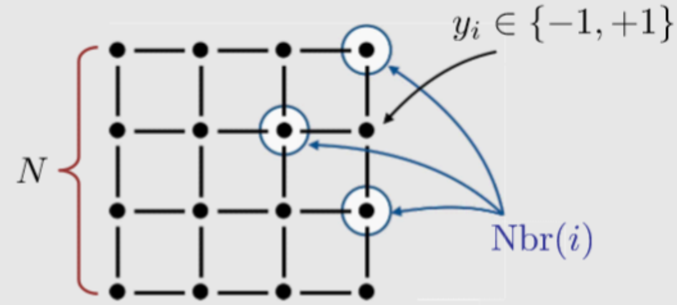
$$= \mathcal{KL}\left[q(z) \parallel \widehat{p}(z)\right] + \log Z$$

$$\mathcal{KL}\left[q(z) \parallel \widehat{p}(z)\right] \to \min_z$$

varinf4

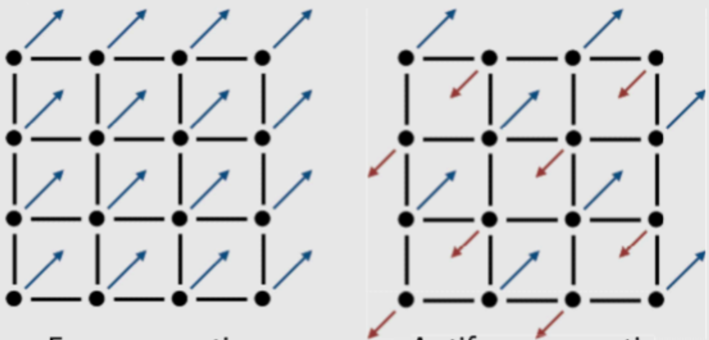
Example: Ising model

Ising model



$$p(y) \propto \exp(\underbrace{\frac{1}{2} J \sum_i \sum_{j \in \mathrm{Nbr}(i)} y_i y_j + \sum_i b_i y_i}_{\phi(y)})$$

Ising model



Ferromagnetic
$J > 0$

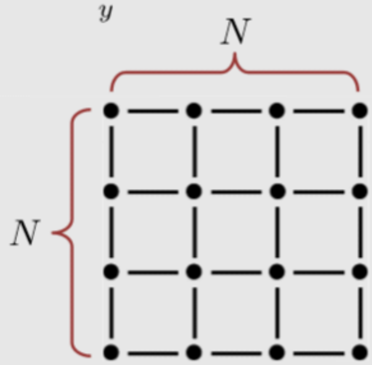
Antiferromagnetic
$J < 0$

$$p(y) \propto \underbrace{\exp(\frac{1}{2} J \sum_i \sum_{j \in \mathrm{Nbr}(i)} y_i y_j + \sum_i b_i y_i)}_{\phi(y)}$$

Normalization constant

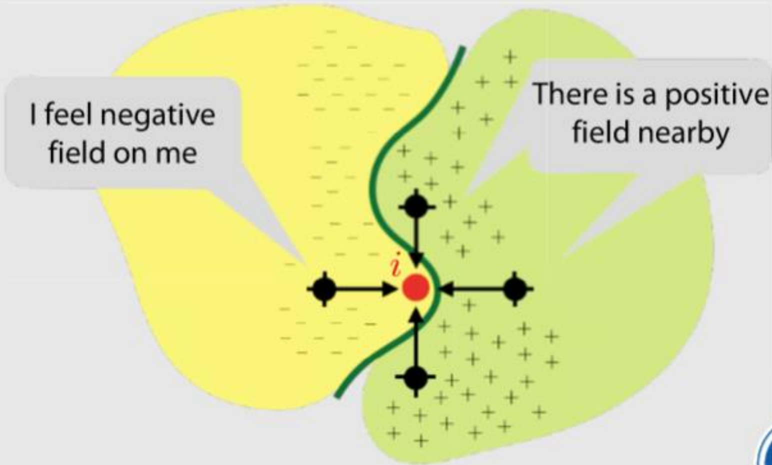
$$p(y) = \frac{1}{Z}\phi(y)$$

$$Z = \sum_y \phi(y) \longleftarrow 2^{N^2} \text{ terms}$$



Mean field

$$p(y) \approx q(y) = \prod_i q_i(y_i)$$



Технический слайд (5 минут на доску)

$$
\begin{aligned}
&\log q_i(y_i) = \mathbb{E}_{y \setminus y_i} p(y) + \text{const} \\
&= \mathbb{E}_{y \setminus y_i} J \sum_{j \in \mathrm{Nbr}(i)} y_i y_j + b_i y_i + \text{const} \\
&= J \sum_{j \in \mathrm{Nbr}(i)} y_i \mathbb{E} y_j + b_i y_i + \text{const} \\
&= J \sum_{j \in \mathrm{Nbr}(i)} y_i \mu_j + b_i y_i + \text{const} \\
&= y_i \overbrace{\left(J \sum_{j \in \mathrm{Nbr}(i)} \mu_j + b_i\right)}^{M} + \text{const} \\
&= M y_i + \text{const}
\end{aligned}
$$


Технический слайд
$q_i(y_i) = \text{const} \cdot e^{My_i}$
$q_i(+1) + q_i(-1) = \text{const}(e^{M} + e^{-M}) = 1$
$q_i(+1) = \frac{e^{M}}{e^{M} + e^{-M}} = \sigma(2M)$
$q_i(-1) = \frac{e^{-M}}{e^{M} + e^{-M}} = 1 - \sigma(2M)$
$M = (J \sum_{j \in \mathrm{Nbr}(i)} \mu_j + b_i)$

Технический слайд

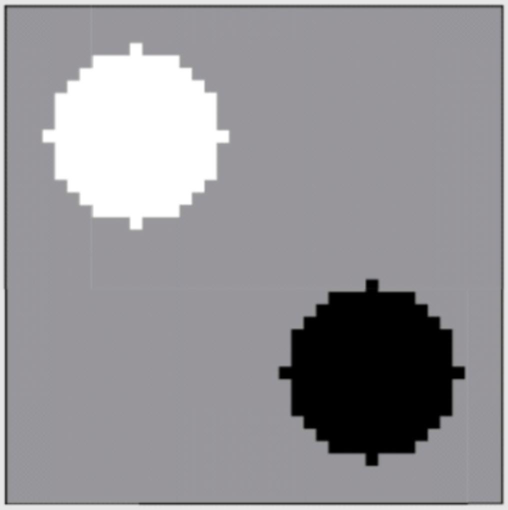
$$q_k(y_k) \propto \exp\Big(J y_k \sum_{j \in Nb(k)} \mu_j\Big) = \exp(y_k M), \quad M = J \sum_{j \in Nb(k)} \mu_j$$

$$q_k(+1) = \frac{e^{M}}{e^{M} + e^{-M}} = \frac{1}{1+e^{-2M}} = \sigma(2M)$$

$$\mu_k = q_k(+1) - q_k(-1) = \frac{1}{1+e^{-2M}} - \frac{e^{-2M}}{1+e^{-2M}} = \frac{1-e^{-2M}}{1+e^{-2M}} = \frac{e^{M}-e^{-M}}{e^{M}+e^{-M}} = \tanh(M)$$

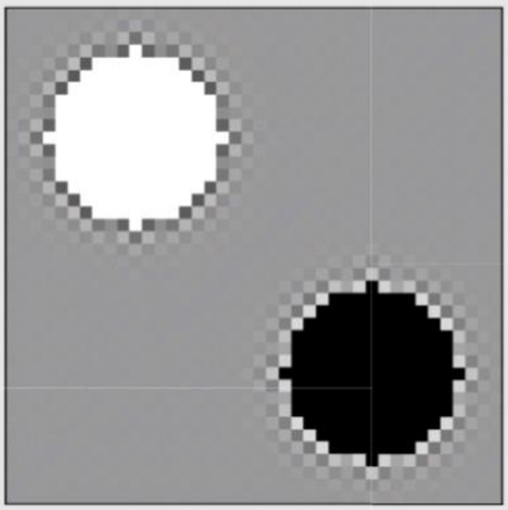
Example

$J = 0$



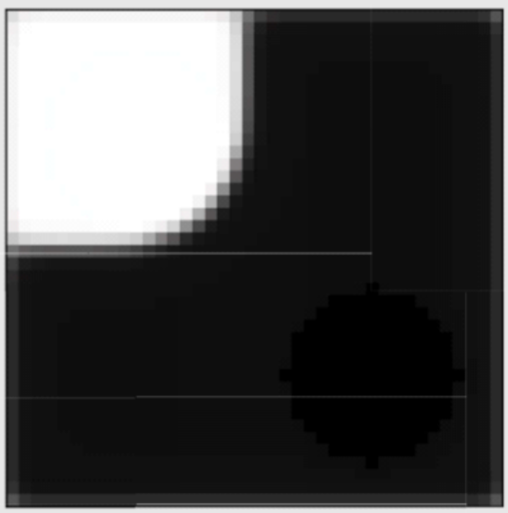
Example

$J = -0.05$

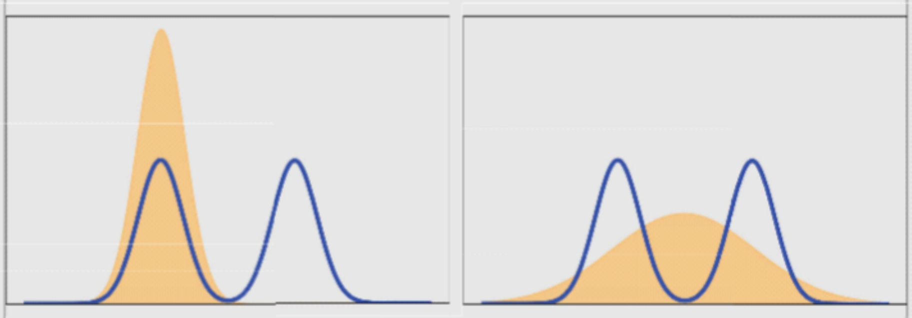


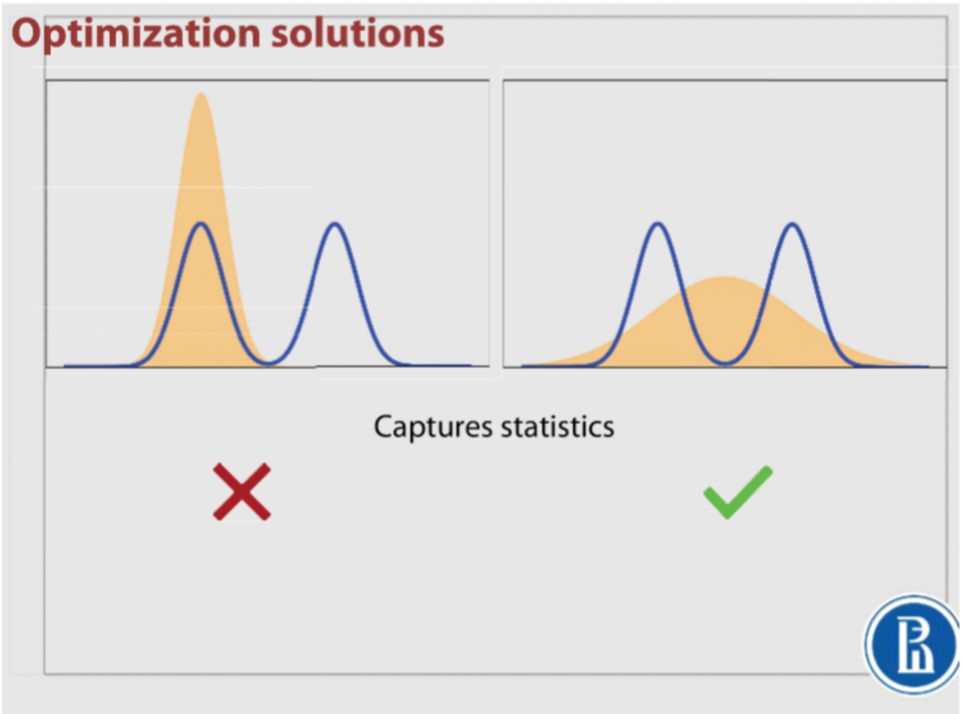
Example

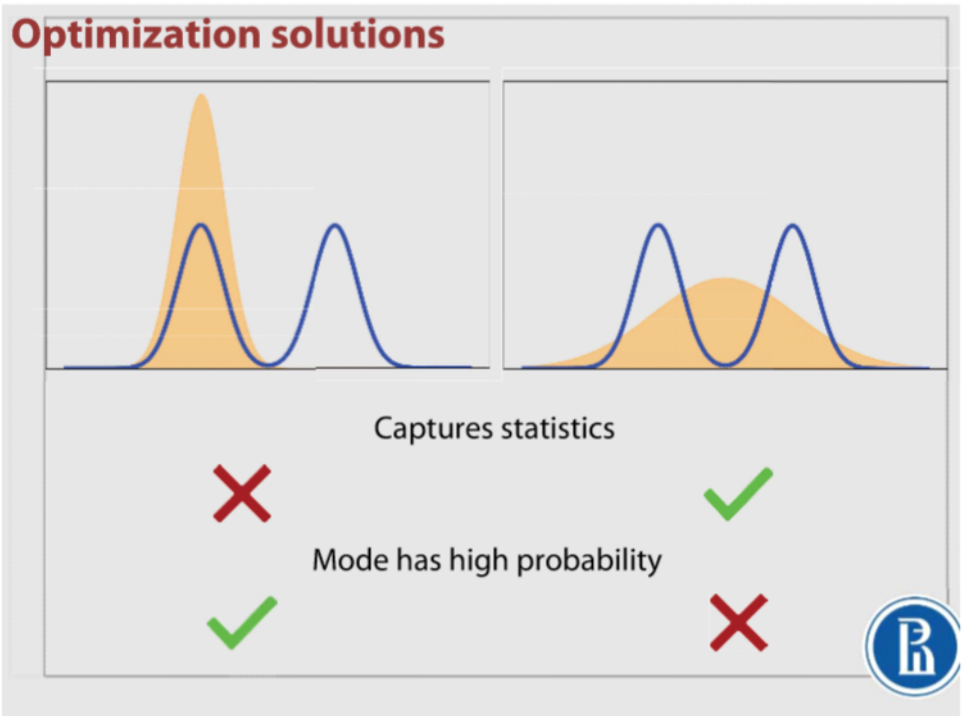
$J = 0.1$

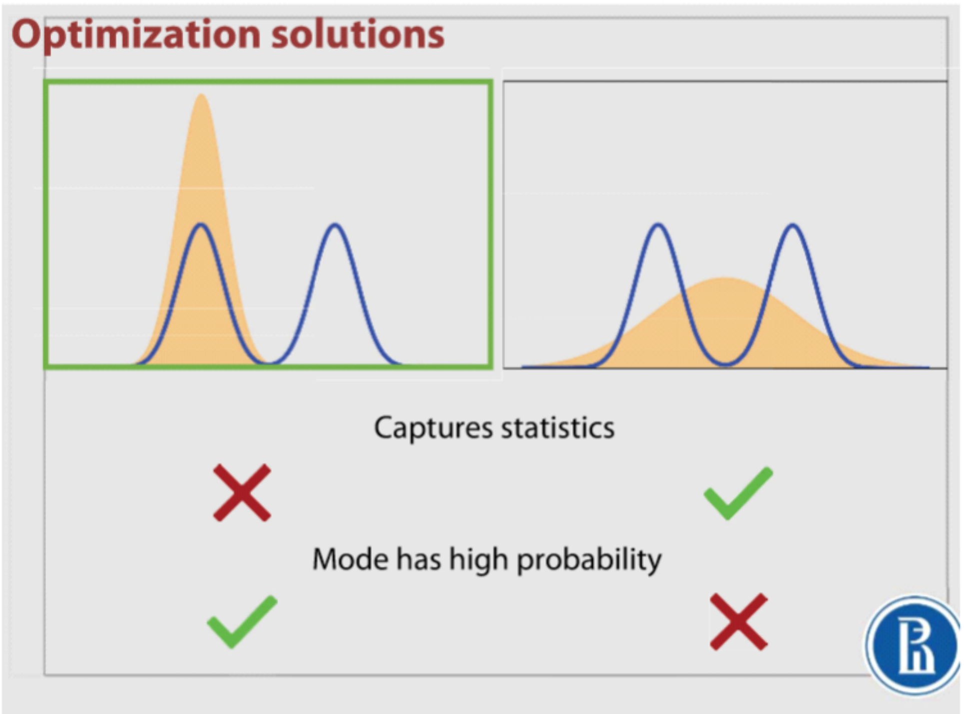


Optimization solutions

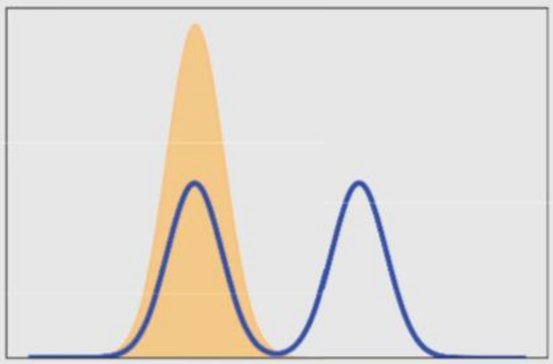



Optimization solutions
Captures statistics


Optimization solutions
Captures statistics
Mode has high probability

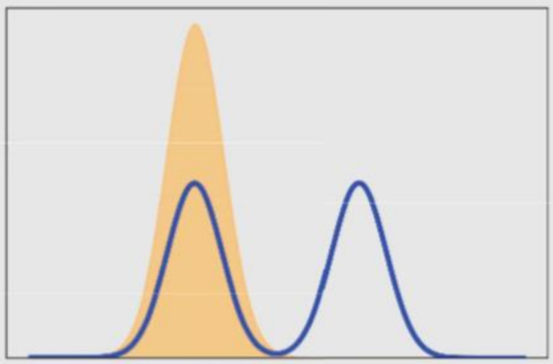

Optimization solutions
Captures statistics
Mode has high probability

Optimization solutions



$$\mathcal{KL}(q \,\|\, p^*) = \int q(z) \log \frac{q(z)}{p^*(z)} dz$$

Optimization solutions

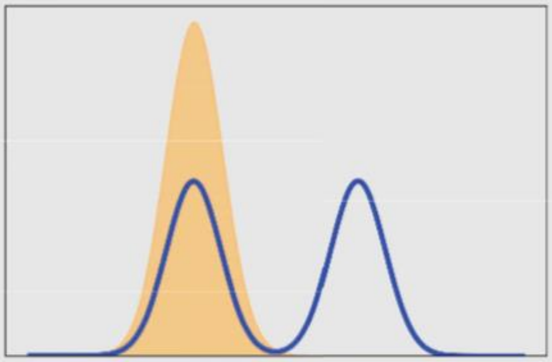


$0 \neq$

$$\mathcal{KL}(q \,\|\, p^*) = \int q(z) \log \frac{q(z)}{p^*(z)} dz$$

$0 =$

Optimization solutions



$$\mathcal{KL}(q \,\|\, p^*) = \int q(z) \log \frac{\overset{0 \neq}{q(z)}}{\underset{0 =}{p^*(z)}} dz = +\infty$$