1. Explicit

$$p_\phi(w)$$
$$w = g(\phi, \varepsilon), \varepsilon \sim p(\varepsilon)$$

   (a) NF – inefficient param
   (b) Neural ODE – slow solvers
   (c) too simple

2. Implicit

$$w = g(\phi, \varepsilon), \varepsilon \sim p(\varepsilon)$$

   (a) GAN-like obj.
   (b) efficient sampling
   (c) biased

   - Adversarial VB $\quad \log \frac{q(w)}{p(w)} \approx D$
   - KIVI (kernel implicit vi) – костыльно
   - Denoising AE

   Попробуем скрестить, так чтобы вывод был проще

3. Semi-implicit

$$p_\phi(w) = \int p_\phi(w \mid \psi) p_\phi(\psi) d\psi$$

$$\mathcal{L} = \mathbb{E}_{q(w)} \log p(y \mid x, w) + \mathbb{E}_{q(w)} \log p(w) - \mathbb{E}_{q(w)} \log q(w)$$

$$\psi = g_\psi(\phi, \Sigma_\psi)$$
$$w = g_w(\phi, \Sigma_w, \psi)$$

   Проблемы с последними двумя членами в $\mathcal{L}$

   Предполагаем $p_\phi(w \mid \psi)$ — explicit, $p_\phi(\psi)$ — implicit. Можно и наоборот, тогда будет другой метод. В любом случае, $p_\phi(w)$ будет implicit.

   Мы это уже видели в VAE. Там такой декодер.

$$p(z), p(x \mid z)$$
$$p(x) = \int p(x \mid z) p(z) dz$$

   $p(x)$ получается в итоге неявным.

- VAE:

$$p(z) = N(0, I)$$
$$p(x|z) = N(\mu_\phi(z), \sigma^2_\phi(z))$$
$$p(x) \text{ – implicit (semi)}$$

- Gaussian Mixture

Не особо уступает implicit моделям, тк можно $p(x|z)$ любой нейронкой задать, а $p(z)$ — гауссианка. При $\sigma^2_\phi \to 0$ получается в чистом виде implicit, но при этом теряем стабильность, тк теряем гладкость.

Итак, надо разобраться с последними двумя слагаемыми.

1. HVM:

$$q(w) \text{ – SI}$$
$$q(w) = \int q_\phi(w \mid \psi) q_\phi(\theta) d\psi \text{ (где } q_\phi(w \mid \psi) q_\phi(\theta) \text{ – explicit)}$$
$$-\mathbb{E}_{q(w)} \log q(w) \geq ?$$

$$q(\psi \mid w) = \frac{q(w \mid \psi) q(\psi)}{q(w)}$$
$$-\mathbb{E}_{q(w)} \log q(w) = -\mathbb{E}_{q(w)} \log \frac{q(w \mid \psi) q(\psi)}{q(\psi|w)}$$
$$-\mathbb{E}_{q(w)} \log q(w) = -\mathbb{E}_{q(w,\psi)} \log \frac{q(w \mid \psi) q(\psi)}{q(\psi|w)} =$$
$$-\mathbb{E}_{q(w,\psi)} \log q(w, \psi) + \mathbb{E}_{q(w,\psi)} q(\psi|w) =$$
$$-\mathbb{E}_{q(w,\psi)} \log q(w, \psi) + \mathbb{E}_{q(w) q(\psi|w)} q(\psi|w) =$$
$$-\mathbb{E}_{q(w,\psi)} \log q(w, \psi) + \mathbb{E}_{q(w) q(\psi|w)} q(\psi|w) \frac{r_\theta(\psi \mid w)}{r_\theta(\psi \mid w)} \geq$$
$$-\mathbb{E}_{q_\phi(w,\psi)} \log q_\phi(w, \psi) + \mathbb{E}_{q_\phi(w,\psi)} \log r_\theta(\psi \mid w)$$

Можно смотреть как на вложенную процедуру вар. вывода $r_\theta(\psi \mid w) \approx q(\psi \mid w)$. По сути запихнули vae в vae, чтобы можно было делать vae, пока мы делаем vae.

$$\mathcal{L} \geq \mathcal{L}'_q(\phi, \theta) \to \max_{\phi, \theta}$$

$$\mathcal{L}'_q = \mathbb{E}_{q(w)} \log p(y \mid x, w) p(w) - \mathbb{E}_{q(w,\psi)} \log q(w, \psi) + \mathbb{E}_{q(w,\psi)} r_\theta(\psi|w)$$
$$= \mathbb{E}_{q(w)} \log p(y \mid x, w) - \mathbb{E}_{q(w,\psi)} \log \frac{q(w, \psi)}{p(w) r(\psi \mid w)}$$
$$\mathrm{KL}\left(q(w) \| p(w)\right) \leq \mathrm{KL}\left(q(w, \psi) \| p(w) r(\psi \mid w)\right)$$

Разобрались с первым случаем. Теперь давайте разбираться с тем, что будет, когда p тоже будет полунеявным.

2.

$$q(w) - \text{explicit}$$

$$p(w) = \int p(w \mid \zeta)p(\zeta)d\zeta \text{ (где } p(w \mid \zeta)p(\zeta) - \text{explicit)}$$

$$\log p(w) \geq \mathbb{E}_{r(\zeta|w)} \log p(w \mid \zeta) - \mathrm{KL}\left(r(\zeta \mid w) \| p(\zeta)\right) =$$
$$\mathbb{E}_{r_\phi(\zeta|w)} \log \frac{p_\theta(w \mid \zeta)p_\theta(\zeta)}{r_\phi(\zeta \mid w)}$$

$$\mathcal{L}'_p = \mathbb{E}_{q(w)} \log p(y \mid x, w) - \mathbb{E}_{q(w)r(\zeta|w)} \log \frac{q(w)r(\zeta \mid w)}{p(w \mid \zeta)p(\zeta)}$$

3.

$$\begin{aligned} p(w) &= \int p(w \mid \zeta)p(\zeta)d\zeta \text{ (где } p(w \mid \zeta) - \text{expl., } p(\zeta) - \text{impl.)} \\ &= \mathbb{E}_{p(\zeta)} p(w \mid \zeta) \\ &\approx \frac{1}{K} \sum_{k=1}^{K} p(w \mid \zeta^k) \end{aligned}$$

$$\log p(w) \geq ?$$
$$\log p(w) \geq \frac{1}{K} \sum_{k=1}^{K} p(w \mid \zeta^k)$$
$$\mathbb{E}_{\zeta^{1\ldots K} \sim p(\zeta)} \log \frac{1}{K} \sum_{k=1}^{K} p(w \mid \zeta^k) \leq \log \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{\zeta^k \sim p(\zeta)} p(w \mid \zeta^k) = \log p(w)$$

$$\left[\mathbb{E}_{q(w)} \log p(w) \geq \mathbb{E}_{\zeta^{1\ldots K}} \mathbb{E}_{q(w)} \log \frac{1}{K} \sum_{k=1}^{K} p(w \mid \zeta^k) \right]$$

4.

$$q(w) = \int q(w \mid \zeta)q(\zeta)d\zeta \text{ (где } q(w \mid \zeta) - \text{expl., } q(\zeta) - \text{impl.)}$$

$$\log q(w) \to \log \frac{1}{K} \sum_{k=1}^{K} q(w \mid \psi)$$

не работает – не в ту сторону неравенство.

$$\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^0)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)$$
$$w\sim q(w\mid\psi^0)$$
$$q(w\mid\psi^0)>q(w\mid\psi^k)$$

Докажем, что такими изменениями превращаем верхнюю оценку в нижнюю.

$$-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^0)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)=$$
$$-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^i)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)$$

Если у нас есть $a\ a\ a\ a\ a\ a$, то

$$a=\frac{a+a+a+\dots+a}{K+1}$$

$$-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^0)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)=$$
$$-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^i)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)=$$
$$-\frac{1}{K+1}\sum_{i=0}^{K}\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^i)}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^k)=$$
$$-\frac{1}{K+1}\sum_{i=0}^{K}\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^i)}\log q(w\mid\psi^{0\dots K})=$$
$$-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{q(w|\psi^{0\dots K})}\log q(w\mid\psi^{0\dots K})$$

$$-\mathbb{E}_{q(w)}\log q(w)=$$
$$-\mathbb{E}_{\phi^{0\dots K}}\mathbb{E}_{q(w|\psi^{0\dots K})}\log q(w)=$$
$$\mathbb{E}_{\phi^{0\dots K}}\mathbb{E}_{w|\psi^{0\dots K}}\log\frac{q(w\mid\psi^{0\dots K})}{q(w)}\geq 0$$
$$\mathbb{E}_{\psi^{0\dots K}}\mathrm{KL}(q(w\mid\psi^{0\dots})\|q(w))\geq 0$$

$$-\mathbb{E}_{q(w)}\log q(w)\geq-\mathbb{E}_{\psi^{0\dots K}}\mathbb{E}_{w|\psi^0}\log\frac{1}{K+1}\sum_{k=0}^{K}q(w\mid\psi^K)$$