$$p(X, Z \mid \Theta) = p(X \mid Z, \Theta)\, p(Z \mid \Theta)$$

$$X = (x_i)_{i=1}^{n}, \quad p(X \mid \Theta) = \prod_n p(x_n \mid \Theta) \to \max_\Theta$$

стохастический варвывод

$$\ln p(X \mid \Theta) = \sum_n \ln p(x_n \mid \Theta) = \sum_n \int q(z \mid x_n) \ln \frac{p(x_n, z \mid \Theta)}{q(z \mid x_n)} dz +$$

$$+ \int q(z \mid x_n) \ln \frac{q(z \mid x_n)}{p(z \mid x_n, \Theta)} dz \geq \sum_n \int q(z \mid x_n) \ln \frac{p(x_n, z \mid \Theta)}{q(z \mid x_n)} dz \to \max_{\Theta,\, q(z \mid x)}$$

$$Q = \{ q(z \mid x, \varphi) \mid \varphi \in \Phi \}$$

$q(z \mid x, \varphi)$ будем параметризовать глубинной нейросетью

и $\to \max$ по Θ, φ

несмещённые оценки

$f(x)$	stoch grad
$\sum_{i=1}^{n} f_i(x)$	$n \nabla f_i(x), \quad i \sim \mathrm{Unif}\{1 \ldots n\}$
$\int p(y)\, h(x, y)\, dy$	$\frac{\partial}{\partial x} \int p(y) h(x,y) dy = \int p(y) \frac{\partial}{\partial x} h(x,y) dy \approx \frac{\partial}{\partial x} h(x, \hat{y}) \underbrace{\int p(y) dy}_{1} = \frac{\partial}{\partial x} h(x, \hat{y}), \quad \hat{y} \sim p(y)$ МСМС
$\sum_{i=1}^{n} \int p(y)\, h(x_i, y)\, dy$	$n \frac{\partial}{\partial x} h(x_i, \hat{y}), \quad i \sim \mathrm{Unif}\{1 \ldots n\}, \ \hat{y} \sim p(y)$
$\int p(y \mid x)\, h(x, y)\, dy$	$\frac{\partial}{\partial x} \int p(y \mid x) h(x,y) dy = \int \frac{\partial}{\partial x} [p(y \mid x)] h(x,y) dy + \int p(y \mid x) \frac{\partial}{\partial x} h(x,y) dy = \int h(x,y) p(y \mid x) \frac{\partial \ln p(y \mid x)}{\partial x} dy + \int p(y \mid x) \frac{\partial}{\partial x} h(x,y) dy \approx h(x, \hat{y}) \frac{\partial}{\partial x} \ln p(\hat{y} \mid x) + \frac{\partial}{\partial x} h(x, \hat{y}), \quad \hat{y} \sim p(y \mid x)$

log derivative trick

$$\frac{df(x)}{dx} = f(x)\,\frac{d\ln f(x)}{dx}$$

стохастические градиенты

$$\int h(x,y)\frac{\partial}{\partial x}p(y|x)\,dy = \int h(x,y)\,p(y|x)\,\frac{\partial \ln p(y|x)}{\partial x}\,dy$$

$$\sum_{i=1}^{n}\mathcal{L}_i(\theta,\varphi) = \sum_{i=1}^{n}\int q(z|x_i,\varphi)\ln\frac{p(x_i,z|\theta)}{q(z|x_i,\varphi)}\,dz$$

$$\frac{\partial \mathcal{L}_\varphi}{\partial\theta} \approx n\cdot\frac{\partial}{\partial\theta}\ln p(x_i,\hat{z}|\theta),\quad \hat{z}\sim q(z|x_i,\varphi)$$

$$i\sim \mathrm{Unif}\{1\ldots n\}$$

$$\frac{\partial\mathcal{L}}{\partial\varphi}\approx n\Big[\underbrace{\ln\frac{p(x_i,\hat{z}|\theta)}{q(\hat{z}|x_i,\varphi)}}_{\to-\infty}\frac{\partial}{\partial\varphi}\ln q(\hat{z}|x_i,\varphi) -$$

$$-\frac{\partial}{\partial\varphi}\ln q(\hat{z}|x_i,\varphi)\Big]$$

большая дисперсия

$$\int q(z|x_i,\varphi)\ln\frac{p(x_i,z|\theta)}{q(z|x_i,\varphi)}\,\frac{\partial \ln q(z|x,\varphi)}{\partial\varphi}\,dz +$$

$$+\underbrace{\int B(x,\varphi)\, q(z|x_i,\varphi)\,\frac{\partial \ln q(z|x_i,\varphi)}{\partial\varphi}\,dz}_{0} =$$

$$=\int q(z|x_i,\varphi)\frac{\partial}{\partial\varphi}\ln q(z|x,\varphi)\Big[\ln\frac{p(x_i,z|\theta)}{q(z|x_i,\varphi)}+B(x,\varphi)\Big]dz\approx$$

$$\approx\frac{\partial}{\partial\varphi}\ln q(\hat{z}|x_i,\varphi)\Big[\ln\frac{p(x_i,\hat{z}|\theta)}{q(\hat{z}|x_i,\varphi)}+\underbrace{B(x,\varphi)}_{\text{Baseline}}\Big]$$

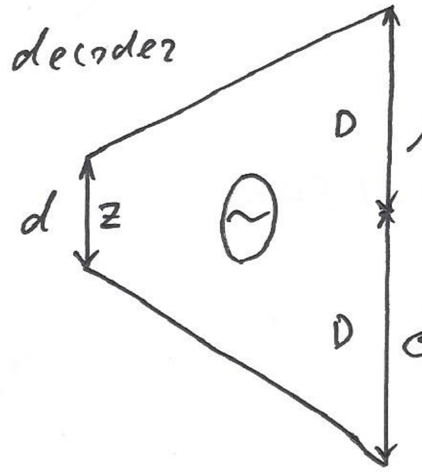
Вариационный автокодировщик

$$p(X,Z|\theta)=\prod_{i=1}^{n} p(x_i|z_i,\theta)\,p(z_i)=\prod_{i=1}^{n} N(x_i|Wz_i+\mu,\sigma^2)\cdot$$

$$\cdot N(z_i|0,I),\quad x_i\in\mathbb{R}^D,\quad z_i\in\mathbb{R}^d \qquad PCA$$

$$\Rightarrow p(X,Z|\theta)=\prod_{i=1}^{n} N(x_i|\mu(z_i,\theta),\Sigma(z_i,\theta))\,N(z_i|0,I)$$

$$p(x|z,\theta)=\prod_{j=1}^{D} N(x_j|\mu_j(z,\theta),\sigma_j^2(z,\theta))$$



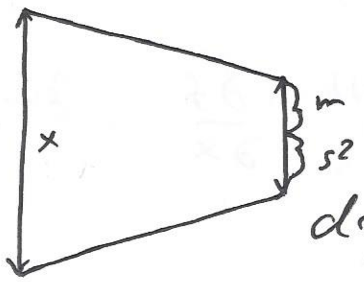
← порождающая нейросеть

$$\mathcal{L}(\theta,\varphi)=\sum_{i=1}^{n}\int q(z|x_i,\varphi)\ln\frac{p(x_i,z|\theta)}{q(z|x_i,\varphi)}dz$$

$$q(z|x,\varphi)=\prod_{j=1}^{d} N(z_j|m_j(x,\varphi),s_j^2(x,\varphi))$$

$$\mathcal{L}(\theta,\varphi)\to\max_{\varphi,\theta}$$

распознающая нейросеть →



encoder

стохастическая оптимизация

doubly stochastic procedure

$$\frac{\partial}{\partial x}\int p(y|x)h(x,y)dy=\left\{\begin{array}{l} y=g(\varepsilon,x),\ \varepsilon\sim p(\varepsilon) \\ p(y|x)dy=p(\varepsilon)d\varepsilon \end{array}\right. \text{репараметризация} \Bigg\}=$$

$$=\frac{\partial}{\partial x}\int p(\varepsilon)h(x,g(\varepsilon,x))d\varepsilon\approx n\frac{d}{dx}h(x,g(\hat\varepsilon,x)) \qquad \hat\varepsilon\sim p(\varepsilon)$$

$$\underline{z_j=s_j(x,\varphi)\varepsilon+m_j(x,\varphi)},\quad \varepsilon\sim N(\varepsilon|0,1)$$

$$\frac{\partial\mathcal{L}}{\partial\varphi}=\frac{\partial}{\partial\varphi}\sum_{i=1}^{n}\int q(\varepsilon)\ln\frac{p(x_i,g(\varepsilon,\varphi,x_i)|\theta)}{q(\underset{g(\varepsilon,\varphi,x_i)}{\cancel{z}}|x_i,\varphi)}d\varepsilon\approx$$

$$\approx n\frac{\partial}{\partial\varphi}\ln\frac{p(x_i,g(\hat\varepsilon,\varphi,x_i)|\theta)}{q(g(\hat\varepsilon,\varphi,x_i)|x_i,\varphi)} \qquad \begin{array}{l}\hat\varepsilon\sim N(\varepsilon|0,I)\\ i\sim\text{Uniform}\{1\ldots n\}\end{array}$$

Stochastic Computational Graph

$x \xrightarrow{f} \circ \xrightarrow{g} \circ \xrightarrow{h} \circ \longrightarrow t$, $t = h(g(f(x)))$

$$\frac{\partial t}{\partial x} = \frac{\partial h}{\partial g}\frac{\partial g}{\partial f}\frac{\partial f}{\partial x}$$

$x \longrightarrow \underset{f}{\circ} \longrightarrow \circ \xrightarrow{\cancel{g}h} \circ \longrightarrow t$

$z \sim p(z \mid f(x))$

$t = h(z)$, $z \sim p(z \mid f(x))$

$$\frac{\partial \mathbb{E}_z t}{\partial x} = \frac{\partial}{\partial x}\int p(z \mid f(x))\, h(z)\, dz$$

1) log derivative trick

$$\frac{\partial \mathbb{E} z}{\partial x} = \underbrace{h(\hat{z})}_{\text{монтек-о}} \frac{\partial \ln p(\hat{z} \mid f(x))}{\partial f}\frac{\partial f}{\partial x} \qquad \hat{z} \sim p(z \mid f(x))$$

2) Reparametrizational trick

$z = g(\varepsilon, f(x))$, $\varepsilon \sim p(\varepsilon)$

$$\frac{\partial \mathbb{E} t}{\partial x} \approx \frac{\partial h}{\partial g}\frac{\partial g(\hat{\varepsilon}, f(x))}{\partial f}\cdot\frac{\partial f}{\partial x}, \quad \hat{\varepsilon} \sim p(\varepsilon)$$

<u>25.11.16 сем</u>

$z = g(\varepsilon, \varphi)$, $\varepsilon \sim q(\varepsilon)$

1. масштаб $\cdot\, \varepsilon$ + сдвиг
2. $F(y) = \mathbb{P}(z < y)$, $\exists F^{-1}$, $z = F^{-1}(u)$, $u \sim U[0,1]$
3. преобразование сл. в. [$y \sim \log Norm$]

Gompertz

$f(z; \eta, b) = b\eta\, e^{\eta + bz} e^{-\eta e^{bz}}$ $u = F(t)$, $t = F^{-1}(u) = g(u; b, \eta)$.

$z \sim U[a, b]$, $g = (b-a)u + a$

$z \sim G(z \mid n, \lambda)$, $n \in \mathbb{Z}$, $\lambda \in \mathbb{R}$ $\quad \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} = \frac{\lambda^n}{n!} x^{n-1} e^{-\lambda x}$

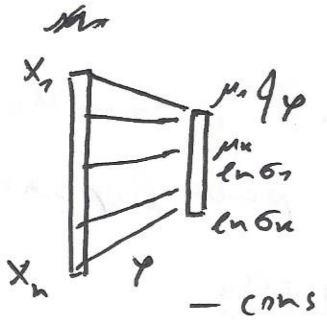
$n = 1$: $\lambda e^{-\lambda x}$, $g = \lambda x$

$$G(n,\lambda) = \sum_{i=1}^{n} G(1,\lambda), \quad J = \frac{1}{\lambda}\sum_{i=1}^{n} J_i$$

$$p_\theta(x|z)\,p_\theta(z)$$

$$\ln p_\theta(x) \geqslant \mathcal{L}(\theta,\varphi,x) = -\mathbb{D}_{KL}\left(q_\varphi(z|x)\,\|\,p(z)\right) +$$

$$+\ \mathbb{E}_{q_\varphi(z|x)} \ln p_\theta(x|z) \xrightarrow[q,\theta]{} \max$$



$$-\|x-\mu_2(z)\|_2^2 + \text{const}$$

const

$$-\text{const}\,\|x-\mu_2(\mu_1(z))\|^2 + \text{const}$$

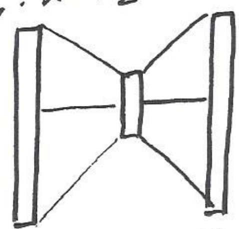
$$p_\theta(x|z) = \mathcal{N}(x\,|\,\mu_2(z), I)$$

$$q_\varphi(z|x) = \delta(z-\mu_1(x))$$

$$\mathbb{E}_{q_\varphi(z|x)} \ln p_\theta(x|z) = \mathbb{E}_{q_\varphi(z|x)} \ln \mathcal{N}(x\,|\,\mu_2(z), I) =$$

$$= \mathbb{E}_{\delta(z-\mu_1(x))} \ln \mathcal{N}(x\,|\,\mu_2(z), I) = \ln \mathcal{N}(x\,|\,\mu_2(\mu_1(z)), I)$$

$\mu_1 : x \to z$



$\mu_2 : z \to x$

$$-c\,\|x-\mu_2(\mu_1(z))\|^2 + c \xrightarrow[q,\theta]{} \max$$

$$\mathbb{D}_{KL}\left(\delta(z-\mu_1(x))\,\|\,p(z)\right) = c + p(\mu_1(x))$$

$$\mu_1 = \mu_1(x;\varphi) \qquad \underline{p(z)=1}$$

несобственный приор: $p(\theta|x) = \dfrac{p(x|\theta)\,p(\theta)}{\int p(x|\theta)\,p(\theta)\,d\theta} \Rightarrow$

$\Rightarrow \dfrac{p(x|\theta)\,\pi(\theta)}{\int p(x|\theta)\,\pi(\theta)\,d\theta}$, причём $\int p(x|\theta)\,\pi(\theta)\,d\theta\ \exists$
если $\pi(\theta)$ – несобственный приор.

$$\pi(\theta) = 1 \qquad p(\theta|x) = \frac{p(x|\theta)}{\int p(x|\theta)\,d\theta}$$

$$\ln p(x) \cong \mathcal{L}_k(q,\theta) \geqslant \mathcal{L}$$

$$\mathcal{L}_k = \mathbb{E}_{z_1 \dots z_k \sim q_\varphi(z|x)} \ln\left(\frac{1}{k}\sum_{i=1}^{k}\frac{p_\theta(x, z_i)}{q_\varphi(z_i|x)}\right) \underset{\text{Йенсен}}{\leqslant} \ln p(x)$$

$$\mathcal{L}_k \geqslant \mathcal{L}_m \quad , \; k \geqslant m$$

$$\{1 \dots k\} \; , \; I = \{i_1 \dots i_m\} \; , \; a_1 \dots a_k$$

$\exists$ C_k^m множеств I равновероятних

$$\mathbb{E}_{I=\{i_1 \dots i_m\}} \frac{a_{i_1} + \dots + a_{i_m}}{m} = \frac{a_1 + \dots + a_k}{k}$$

$$\mathcal{L}_k = \mathbb{E}_{z_1 \dots z_k} \ln\left(\frac{1}{k}\sum_{i=1}^{k}\frac{p(x, z_i)}{q_\varphi(z_i|x)}\right) \geqslant$$

$$\geqslant \mathbb{E}_{z_1 \dots z_k} \cancel{\ln} \, \mathbb{E}_{\{i_1 \dots i_m\}} \ln\left(\frac{1}{m}\sum_{j=1}^{m}\frac{p(x, z_j)}{q_\varphi(z_j|x)}\right) =$$

$$= \mathbb{E}_{\{i_1 \dots i_m\}} \mathbb{E}_{z_{i_1} \dots z_{i_m}} \ln\left(\frac{1}{m}\sum_{j=1}^{m}\frac{p(x, z_j)}{q_\varphi(z_j|x)}\right)$$