Probabilistic Graphical Models

Lectures 30

Variational Autoencoders

Generative Models

K. N. Toosi
University of Technology

Karras, Tero, et al "A style-based generator architecture for generative adversarial networks." CVPR 2019.

Review Variational Inference

$$x^1, x^2, \dots, x^m$$

$$P_\theta(x) = \int P_\theta(x, z)\, dz$$

$$= \int P_\theta(x|z)\, p(z)\, dz$$

log-likelihood

$$\ell\ell(\theta) = \sum_{i=1}^{m} \log P_\theta(x^i) = \sum_{i=1}^{m} \log \int P_\theta(x^i; z)\, dz$$

Hard!

ELBO

$$\mathcal{L}(q_i, x^i; \theta) = \sum q_i(z|x^i) \log \frac{P_\theta(x^i; z)}{q_i(z|x^i)}$$

$$= \log P_\theta(x^i) - KL\left(q_i(z|x^i) \,\|\, P_\theta(z|x^i)\right)$$

$$\max_{\theta} \ell\ell(\theta) \Longrightarrow \max_{\theta, q_1, q_2, \dots, q_m} \sum_{i=1}^{m} \mathcal{L}(q_i, x^i; \theta)$$

$$= \max_{\theta} \sum_{i=1}^{m} \log P_\theta(x^i)$$

Review Variational Inference

start from some $\theta, q_1, q_2, \ldots q_m$

Loop

$$\nabla_\theta = \frac{\partial}{\partial\theta} \sum_{i=1}^{m} \mathcal{L}(q_i, x^i; \theta)$$

$$\theta \leftarrow \theta + \lambda \nabla_\theta$$

for $i = 1 \ldots m$:

$$q_i \leftarrow \underset{q}{\operatorname{argmax}} \, \mathcal{L}(q_i, x^i; \theta)$$

Variational Inference with large datasets

K. N. Toosi
University of Technology

what if we have millions of data?

solution 1 $\quad q_i(z|x^i) \longrightarrow q(z)$ ✗

solution 2: $\quad q_i(z|x^i) \longrightarrow q_\phi(z|x^i)$

$$x^i \longrightarrow \boxed{q_\phi} \longrightarrow z$$

Amortized
Variational
Inference

Amortized Variational Inference

K. N. Toosi

of Technology

Amortized VI

pgm 30 I

start from some θ, ϕ

Loop

$$\nabla_\theta = \frac{\partial}{\partial \theta} \sum_{i=1}^{m} \mathcal{L}(x^i; \phi, \theta)$$

$$\nabla_\phi = \frac{\partial}{\partial \phi} \sum_{i=1}^{m} \mathcal{L}(x^i; \phi, \theta)$$

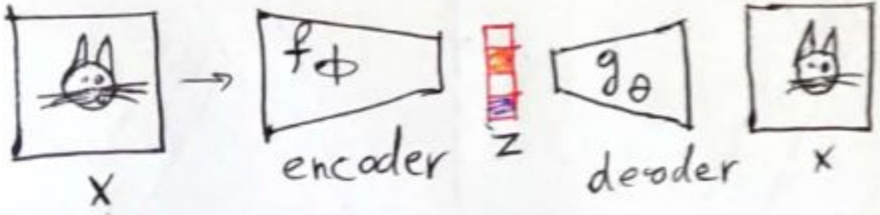
$$\theta \leftarrow \theta + \lambda \nabla_\theta$$

$$\phi \leftarrow \phi + \lambda \nabla_\phi$$

$$\mathcal{L}(x^i; \phi, \theta)$$

$$= \sum_z q_\phi(z|x^i) \log \frac{p_\theta(z, x^i)}{q_\phi(z|x^i)}$$

Autoencoders



$z = f_\phi(x)$

$y = g_\theta(z)$

find ϕ, θ such that $\|y - x\|$ is small

$$\min_{\theta, \phi} \sum_{i=1}^{m} \left\| g_\theta\left(f_\phi(x^i) \right) - x^i \right\|^2$$

Autoencoders' applications

K. N. Toosi
University of Technology

- Data Compression
- Dimensionality Reduction / Manifold Learning

- Unsupervised feature extraction, Representation Learning
- Semi-supervised Learning

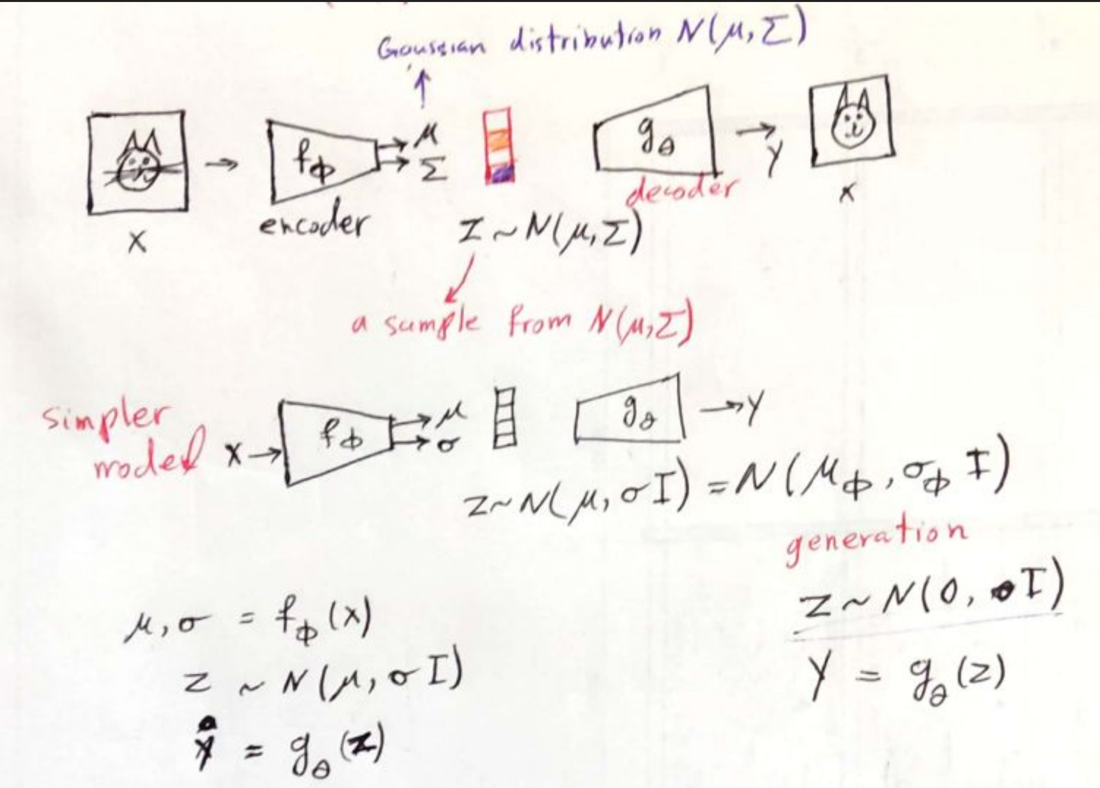
Generating Data Using Autoencoders

choose a random z (e.g. $z \sim \mathcal{N}(0, I)$)

$\hat{x} = g_\theta(z)$

doesn't work becase the distribution of z is unknown $\neq \mathcal{N}(0, I)$

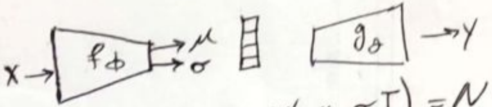
Variational Autoencoders (VAE)



VAE Cost function

K. N. Toosi
University of Technology

simpler model



$z \sim N(\mu, \sigma I) = N(\mu_\phi, \sigma_\phi I)$

generation

$z \sim N(0, I)$

$Y = g_\theta(z)$

$\mu, \sigma = f_\phi(x)$

$z \sim N(\mu, \sigma I)$

$\hat{x} = g_\theta(z)$

$x^1, x^2, \dots, x^m$

$g_\theta(z(\phi))$

$$C(\theta,\phi) = \sum_{i=1}^{m} \alpha \left\| g_\theta\left(z \sim N\left(\mu_\phi(x^i), \sigma_\phi(x^i) I\right)\right) - x^i \right\|^2 + \sum_{i=1}^{m} KL\left(N\left(\mu_\phi(x^i), \sigma_\phi^2(x^i) I\right) \,\|\, N(0, I)\right)$$

How to compute $\frac{\partial}{\partial \theta} C(\theta,\phi)$, $\frac{\partial}{\partial \phi} C(\theta,\phi)$?

How to compute derivatives?

How to compute $\frac{\partial}{\partial \theta} C(\theta, \phi)$, $\frac{\partial}{\partial \phi} C(\theta, \phi)$?

$\mu, \sigma = f_\phi(x)$ ✓

$z \sim N(\mu, \sigma I)$?

$y = g_\theta(z)$ ✓

How to diffentiate from the sampling operation?

Reparameterization Trick

$$z \sim \mathcal{N}(\mu_\phi, \sigma_\phi^2 I)$$

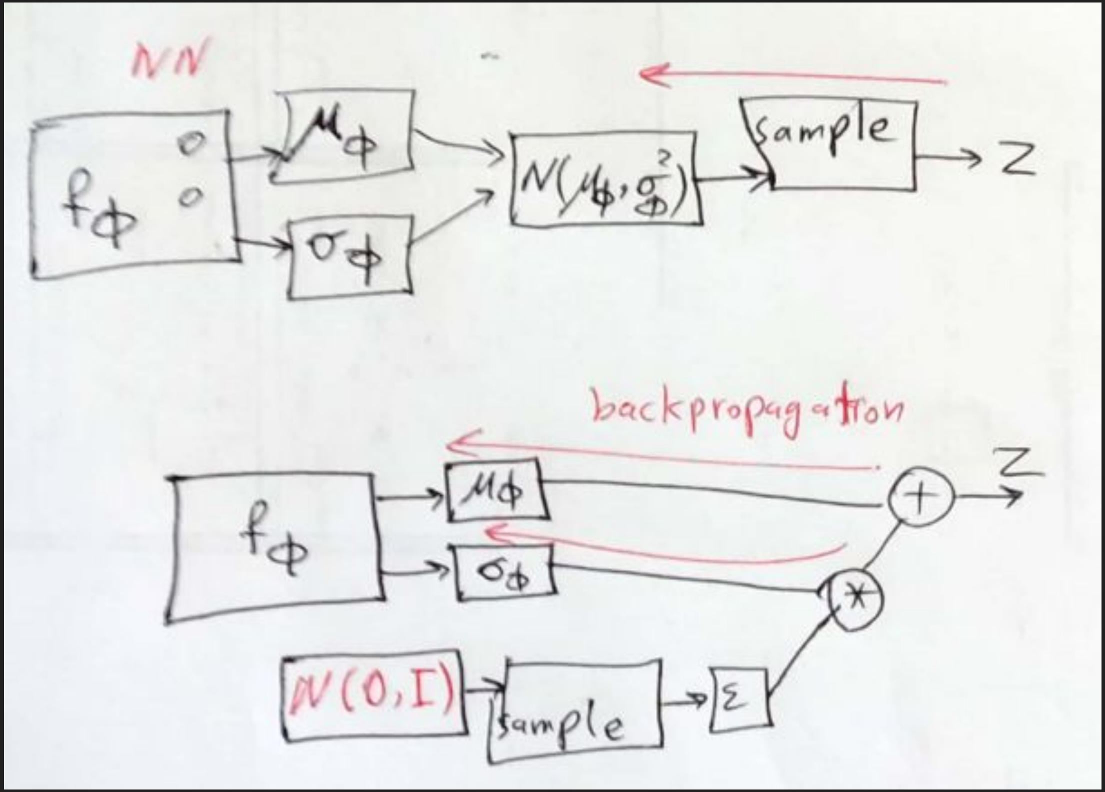
$$\downarrow$$

$$z = \mu_\phi + \varepsilon \, \sigma_\phi$$

$$\varepsilon \sim \mathcal{N}(0, I)$$

$$\boxed{\begin{aligned} &\varepsilon \sim \mathcal{N}(0, I) \\ &\sigma\varepsilon \sim \mathcal{N}(0, \sigma^2 I) \\ &\mu + \sigma\varepsilon \sim \mathcal{N}(\mu, \sigma^2 I) \end{aligned}}$$

Reparameterization Trick



$$\varepsilon \sim N(0, I)$$

$$Z = \mu_\phi + \varepsilon\, \sigma_\phi$$

$$\frac{\partial Z}{\partial \phi} = \frac{\partial \mu_\phi}{\partial \phi} + \varepsilon \frac{\partial \sigma_\phi}{\partial \phi}$$

New Cost Function

K. N. Toosi
University of Technology

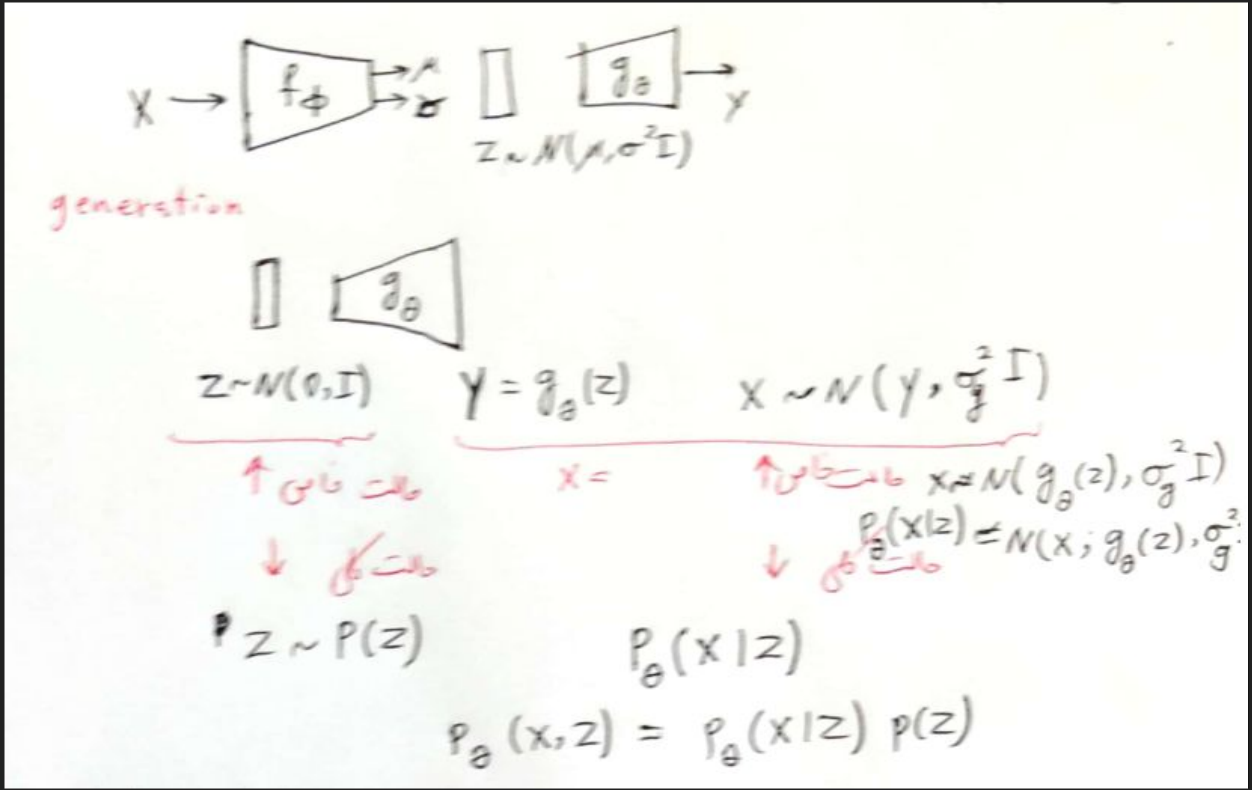
$$C(\theta, \phi) = \sum_{i=1}^{m} \alpha \left\| g_\theta\left(z \sim \mathcal{N}\left(\mu_\phi(x^i), \sigma_\phi(x^i) I\right) - x^i \right\| + \sum_{i=1}^{m} KL\left(\mathcal{N}\left(\mu_\phi(x^i), \sigma_\phi^2(x^i) I\right) \| \mathcal{N}(0, I)\right)$$

$\varepsilon^i \sim \mathcal{N}(0, I)$

$$C(\theta, \phi) = \sum_{i=1}^{m} \alpha \| g_\theta\left(\mu_\phi(x^i) + \varepsilon^i \sigma_\phi(x^i) I\right) + KL\left(\mathcal{N}\left(\mu_\phi(x^i), \sigma_\phi^2(x^i) I\right) \| \mathcal{N}(0, I)\right)$$

Variational Inference View

K. N. Toosi
University of Technology



Variational Inference View

data $x^1, x^2, \cdots, x^m$

$$\max_\theta \; ll(\theta) = \sum_{i=1}^{m} \log P_\theta(x^i) = \sum_{i=1}^{m} \log \int P_\theta(x, z)\, dz$$

log-likelihood

$$= \sum_{i=1}^{m} \log \int P_\theta(x|z)\, p(z)\, dz$$

Hard to maximzied

ELBO (Amotized)

$\rightarrow \boxed{q_\phi} \rightarrow q_\phi(z|x)$ posterior

$$\max_{\phi, \theta} \; \mathcal{L}(\phi, \theta) = \sum_i \int q_\phi(z|x^i) \log \frac{P_\theta(x^i, z)}{q_\phi(z|x^i)}$$

$$= \sum_{i=1}^{m} \int q_\phi(z|x^i) \log \frac{P_\theta(x^i|z)\, p(z)}{q_\phi(z|x^i)}$$

Variational Inference View

K. N. Toosi
University of Technology

Hard to

ELBO (Amotized)

$\rightarrow \boxed{q_\phi} \rightarrow q_\phi(z|x)$ posterior

$$\max_{\phi,\theta} \mathcal{L}(\phi,\theta) = \sum_i \int q_\phi(z|x^i) \log \frac{p_\theta(x^i,z)}{q_\phi(z|x^i)}$$

$$= \sum_{i=1}^{m} \int q_\phi(z|x^i) \log \frac{p_\theta(x^i|z)\, p(z)}{q_\phi(z|x^i)}$$

$$= \sum_{i=1}^{m} \int q_\phi(z|x^i) \log p_\theta(x^i|z) - \int q_\phi(z|x^i) \frac{q_\phi(z|x^i)}{p(z)}$$

$$\sum_{i=1}^{m} E_{q_\phi(z|x^i)}\{\log p_\theta(x^i|z)\} - KL\left(q_\phi(z|x^i) \,\|\, p(z)\right)$$

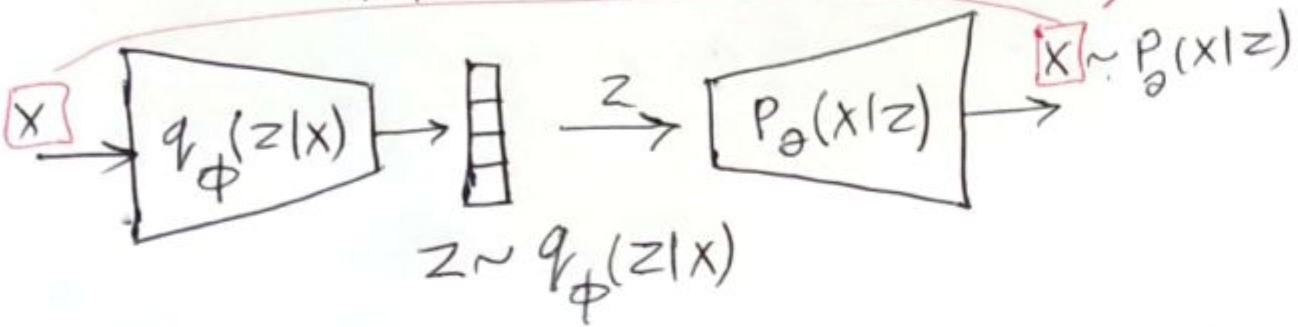
$x \rightarrow \boxed{\text{enc}}$ $q(x|z)$ $\boxed{\text{dec}}$ $p(z|x)$

Variational Inference View

K. N. Toosi
University of Technology

pgm30

$$\mathcal{L}(\phi,\theta) = \sum_{i=1}^{m} E_{q_\phi(z|x^i)}\left\{\log p_\theta(x^i|z)\right\} - \sum_{i=1}^{m} KL\left(q_\phi(z|x^i) \,\|\, p(z)\right)$$



Variational Inference View

K. N. Toosi
University of Technology

حالت خاص

$$q_\phi(z|x) = \mathcal{N}(z;\ \mu_\phi(x^i),\ \sigma_\phi(x^i)^2 I)$$

$\varepsilon \sim \mathcal{N}(0,$

$$z \sim q_\phi(z|x) \Rightarrow z = \mu_\phi(x^i) + \varepsilon\, \sigma_\phi(x^i)$$

$$P_\theta(x|z) = \mathcal{N}(x;\ g_\theta(z),\ \sigma_g I)$$

$$p(z) = \mathcal{N}(0, I)$$

Variational Inference View

$$\mathcal{L}(\phi,\theta) = \sum_{i=1}^{m} E_{q_\phi(z|x^i)} \{ \log N(x; g_\theta(z), \sigma_g I) \}$$

$$- \sum_{i=1}^{m} KL\left(N(\mu_\phi(x^i), \sigma_\phi(x^i)^2) \,\|\, N(0,I) \right)$$

$$= \sum_{i=1}^{m} E_{q_\phi(z|x^i)} \left\{ \log \frac{1}{\sqrt{2\pi}\,\sigma_g^{d/2}} e^{\frac{\|x - g_\theta(z)\|^2}{\sigma_g^{2d}}} \right\} - KL$$

$$= \sum_{i=1}^{m} E_{q_\phi}(z|x^i) \left\{ \cancel{-\frac{d}{2}\log \sigma_g} + \frac{\|x - g_\theta(z)\|^2}{\sigma_g^{2d}} \right\} - KL$$

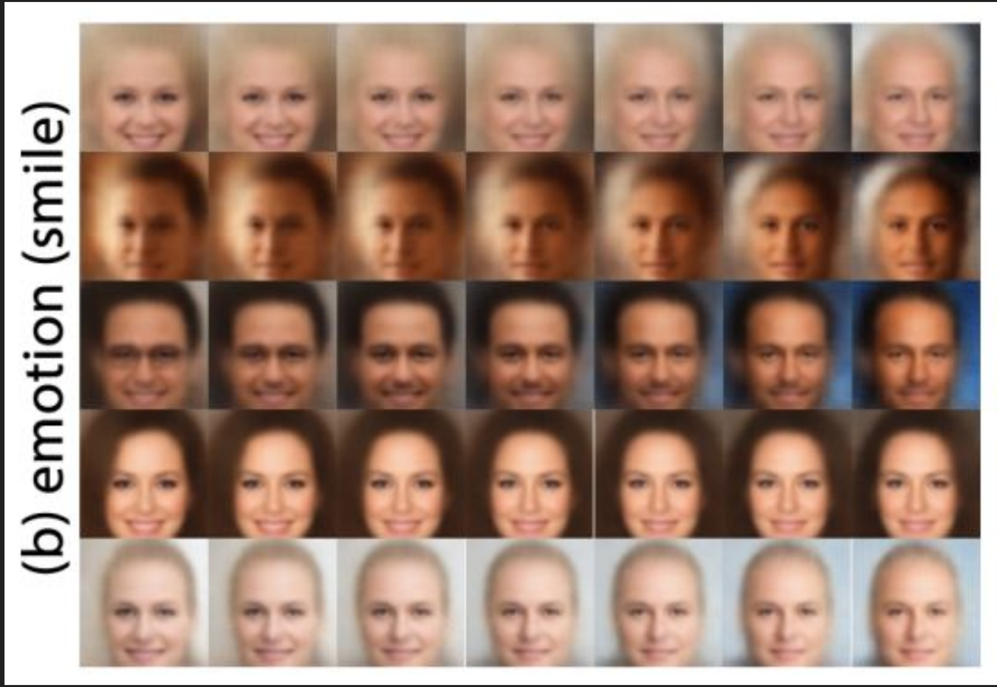
one sample approximation $z^i \sim q_\phi(z|x^i)$

$$\simeq \sum \frac{\|x^i - g_\theta(z^i)\|^2}{\sigma_g^{2d}} + \sum KL$$

beta-VAE

Irina Higgins, et al. "beta-VAE: Learning basic visual concepts with a constrained variational framework." ICLR 2017.

beta-VAE



Irina Higgins, et al. "beta-VAE: Learning basic visual concepts with a constrained variational framework." ICLR 2017.

beta-VAE

K. N. Toosi
University of Technology

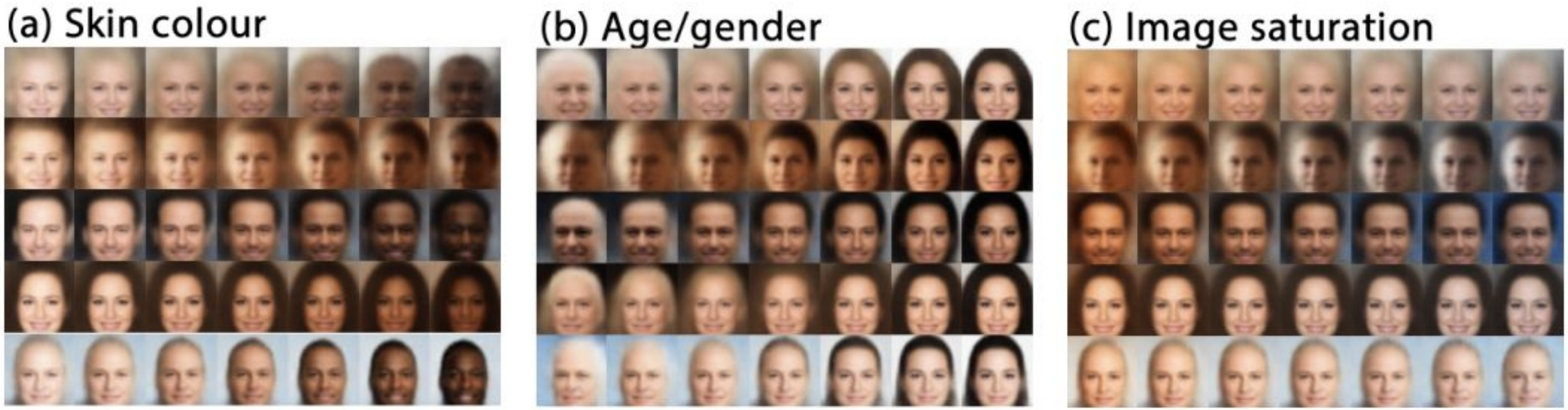


Figure 4: **Latent factors learnt by β-VAE on celebA:** traversal of individual latents demonstrates that β-VAE discovered in an unsupervised manner factors that encode skin colour, transition from an elderly male to younger female, and image saturation.

Irina Higgins, et al. "beta-VAE: Learning basic visual concepts with a constrained variational framework." ICLR 2017.