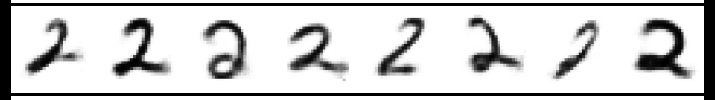
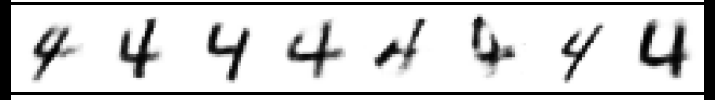
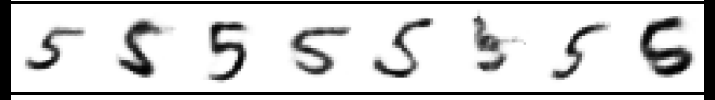
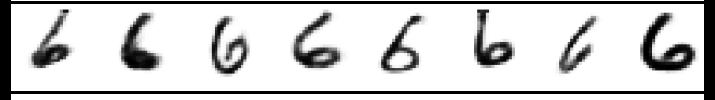
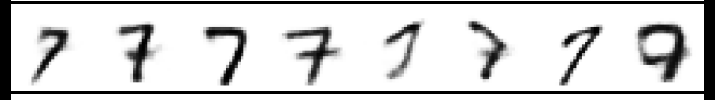
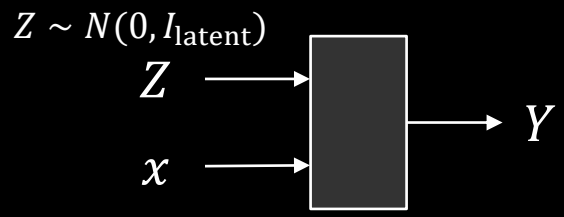
A conditional generator for handwriting

digit x	generated renditions Y of x
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

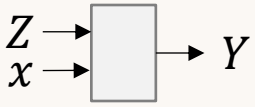
Suggest a generative model behind this.



x is the digit we're asking for
Z is latent noise, to generate variation in Y

$$Y \sim f_\theta(x, Z) + N(0, \sigma^2 I_{\text{data}})$$

Wiring an autoencoder to fit this generator



Given a label x, what is the likelihood of a value y?

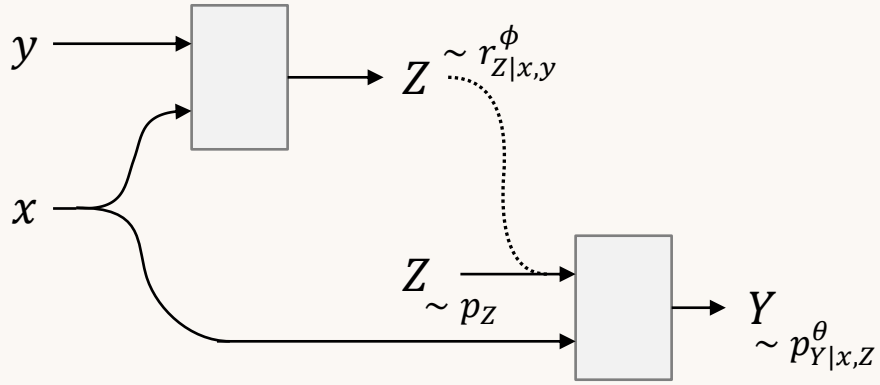
$$p_Y^\theta(y|x) = \mathbb{E}_{Z\sim p_Z}\, p_Y^\theta(y|x,Z)$$

Rewrite this expectation using a sampler $Z \sim r_Z$

$$p_Y^\theta(y|x) = \mathbb{E}_{Z\sim r_Z}\, p_Y^\theta(y|x,Z)\,\frac{p_Z(Z)}{r_Z(Z)}$$

What is the perfect sampler?

- Algebraically, it is $r_Z(z) \propto p_Z(z)\, p_Y^\theta(y|x,z)$, i.e. it depends on both x and y
- Conceptually, it is the posterior distribution $(Z|x,y)$



Give a lower bound on $\log p_Y^\theta(y|x)$

$$\log p_Y^\theta(y|x) = \log \mathbb{E}_{Z\sim r^\phi_{Z|x,y}}\, p_Y^\theta(y|x,Z)\,\frac{p_Z(Z)}{r^\phi_{Z|x,y}(Z)}$$

$$\geq \mathbb{E}_{Z\sim r^\phi_{Z|x,y}} \log\left[p_Y^\theta(y|x,Z)\,\frac{p_Z(Z)}{r^\phi_{Z|x,y}(Z)}\right]$$

$$= \mathbb{E}_{Z\sim r^\phi_{Z|x,y}}\left\{\log p_Y^\theta(y|x,Z) - \mathrm{KL}(r^\phi_{Z|x,y} \parallel p_Z)\right\}$$

$$\mathrm{ELBO}_{x,y}(p^\theta, r^\phi)$$

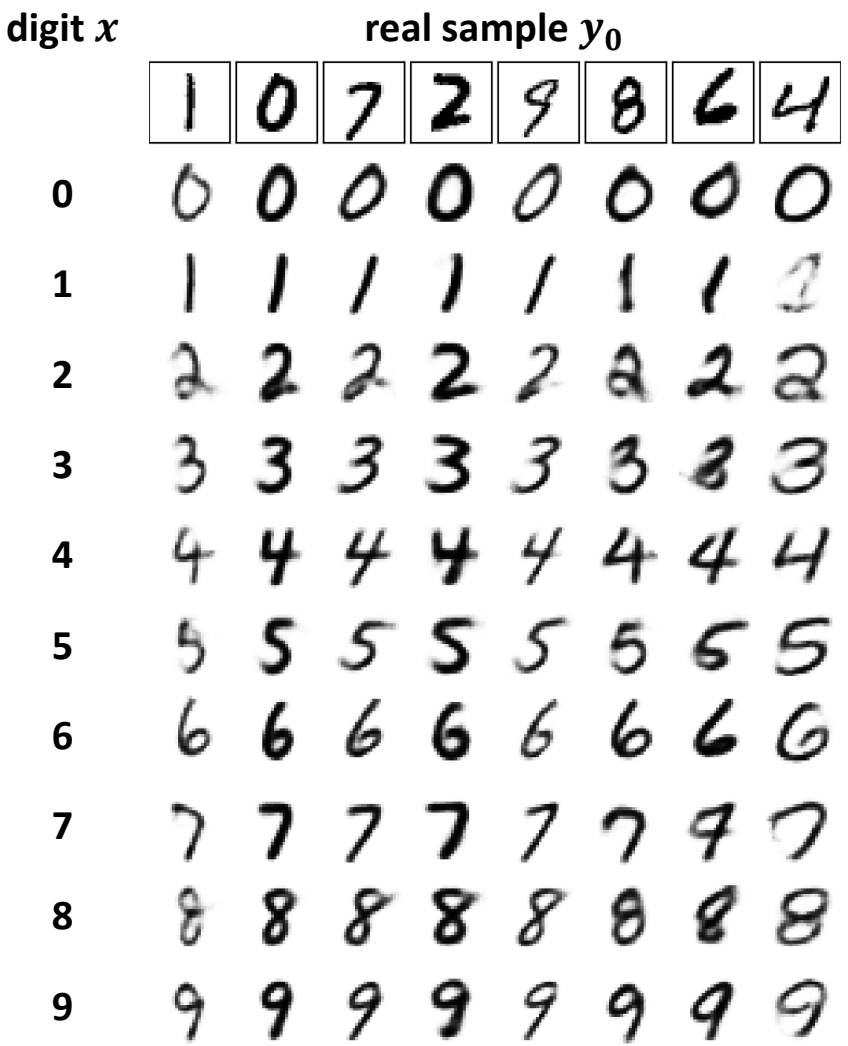
In KL terms,

$$\mathbb{E}_{X\sim q_X}\mathrm{KL}(q_{Y|X} \parallel p^\theta_{Y|X}) \leq \mathrm{const} - \mathbb{E}_{X,Y\sim q_{X,Y}}\mathrm{ELBO}_{X,Y}(p^\theta, r^\phi)$$

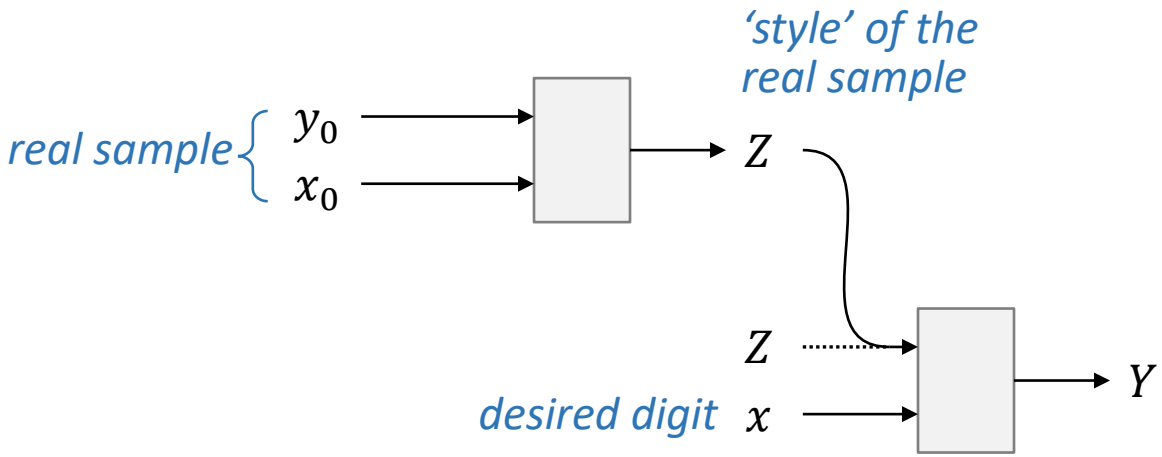
How would you fit the generative model?

$\max_{\theta,\phi} \sum_i \mathrm{ELBO}_{x_i,y_i}(p^\theta, r^\phi)$ and watch out when calculating gradients, as ELBO has an expectation over a ϕ-dependent dist.

Style mimicry



In this picture, each column copies its style from a real image, shown in the top row. How might we achieve this?



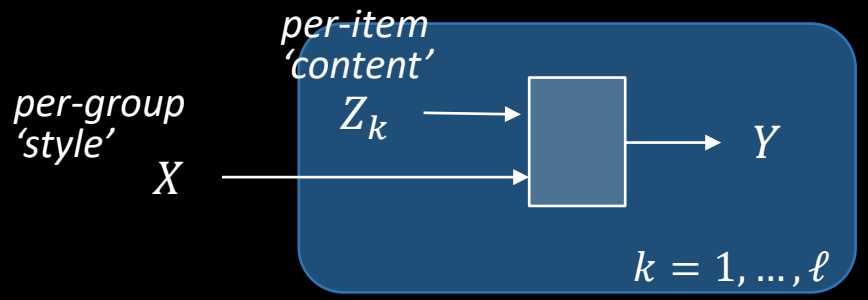
Data that comes in groups

Consider a grouped dataset $\{y_k^{(i)}\}$, where $i \in \{1, \dots, n\}$ indexes the group and $k \in \{1, \dots, \ell_i\}$ indexes the item within a group.



Suggest a generative model for this dataset.

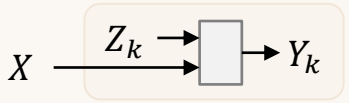
Model for the readings $Y_{1:\ell}$ in a single group:



Assume each group (i.e. each row i) is sampled independently from this single-group model. Then

$$\log p^\theta(\text{dataset}) = \sum_{i=1}^{n} \log p^\theta\left(y_{1:\ell_i}^{(i)}\right)$$

Wiring an autoencoder to fit this grouped model

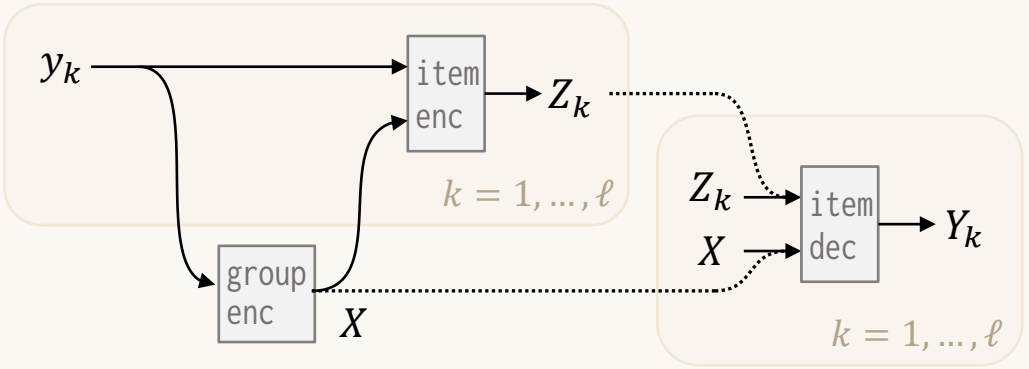


What is the likelihood of a single group, $p^\theta(y_{1:\ell})$?

$$p^\theta(y_{1:\ell}) = \mathbb{E}_{X,Z_{1:\ell}}\, p^\theta(y_{1:\ell}|X, Z_{1:\ell})$$

What would be the perfect sampler for $(X, Z_{1:\ell})$?

The posterior distribution of $(X, Z_{1:\ell})$ conditional on $Y_{1:\ell} = y_{1:\ell}$



Cunning trick.

- By the chain rule the perfect sampler factorizes as $r(x|y_{1:\ell})\, r(z_{1:\ell}|x, y_{1:\ell})$
- Once $X = x$ is fixed the model is ℓ independent plates, so we can encode the items separately with $(Z_k|x, y_k)$

Give a lower bound on $\log p^\theta(y_{1:\ell})$ using an encoder with this structure.

$$\log p^\theta(y_{1:\ell}) = \log \mathbb{E}_{X,Z_{1:\ell}\sim r^\phi}\left[p^\theta(y_{1:\ell}|X, Z_{1:\ell})\,\frac{p(X, Z_{1:\ell})}{r^\phi(X, Z_{1:\ell})}\right] \quad \text{importance sampling}$$

$$\geq \mathbb{E}_{X,Z_{1:\ell}\sim r^\phi}\ \log\left[p^\theta(y_{1:\ell}|X, Z_{1:\ell})\,\frac{p(X, Z_{1:\ell})}{r^\phi(X, Z_{1:\ell})}\right] \quad \text{Jensen}$$

$$= \mathbb{E}_{X,Z_{1:\ell}\sim r^\phi}\left[\sum_{k=1}^{\ell} \log\, p^\theta(y_k|X, Z_k) + \log\frac{p_X(X)}{r^\phi_{X|y_{1:\ell}}(X)} + \sum_{k=1}^{\ell}\log\frac{p_Z(Z_k)}{r^\phi_{Z|y_k,X}(Z_k)}\right]$$

using the factorizations in the encoder and the decoder

$$= \mathbb{E}_{X\sim r^\phi_{X|y_{1:\ell}}} \sum_k \left[\mathbb{E}_{Z\sim r^\phi_{Z|y_k,X}} \log p^\theta(y_k|X, Z) - \mathrm{KL}\left(r^\phi_{Z|y_k,X}\,\|\,p_Z\right)\right] - \mathrm{KL}\left(r^\phi_{X|y_{1:\ell}}\,\|\,p_X\right)$$

splitting the expectation over $(X, Z_{1:\ell})$ into one over X and one over $(Z_{1:\ell}|X)$

How would you fit the generative model?

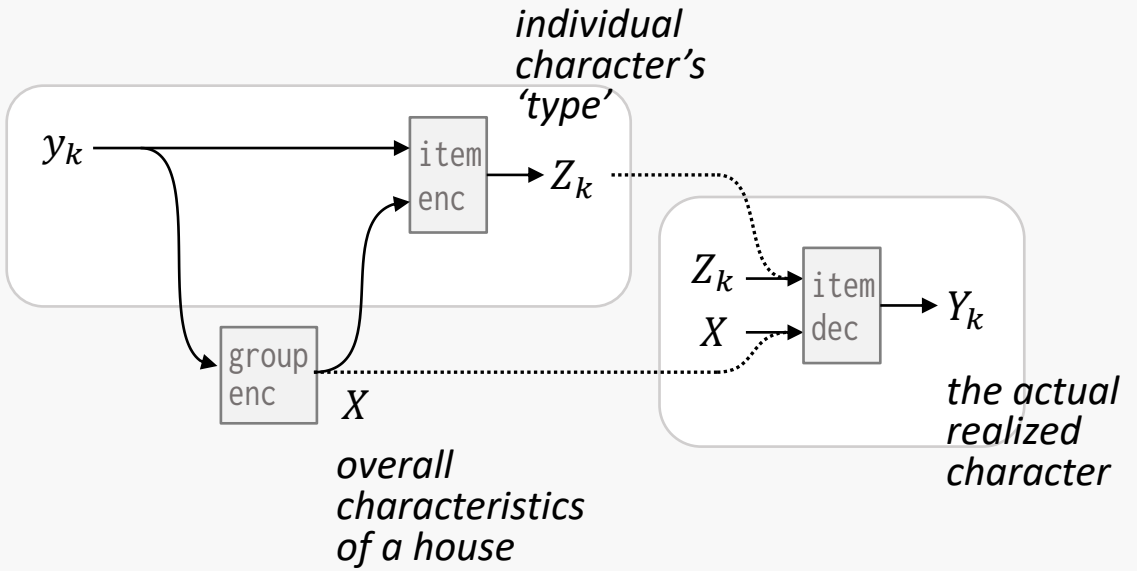
maximize this over θ, ϕ — it's all just things we can approximate with Monte Carlo

The grouped-data model lets us do few-shot generation and translation

house = Gryffindor 

house = Slytherin 

house = Ravenclaw …



Things we might do with this grouped model

Few-shot generation:
a new Slytherin student

1. encode Slytherin to get x_{slyth}
2. generate a fresh $z \sim p_Z$
3. decode (x_{slyth}, z)

Translation:
what would the Harry Potter of Slytherin be?

1. encode Gryffindor and get z_{hp}
2. encode Slytherin to get x_{slyth}
3. decode $(x_{\text{slyth}}, z_{\text{hp}})$

- If we had per-school-house predictor variables x, this would be a conditional generation problem, same as CVAE
- We don't observe x, but we do know that other members of the group share it — so let's infer a 'pseudo-predictor' $x = \text{group_enc}(y_{1:\ell})$



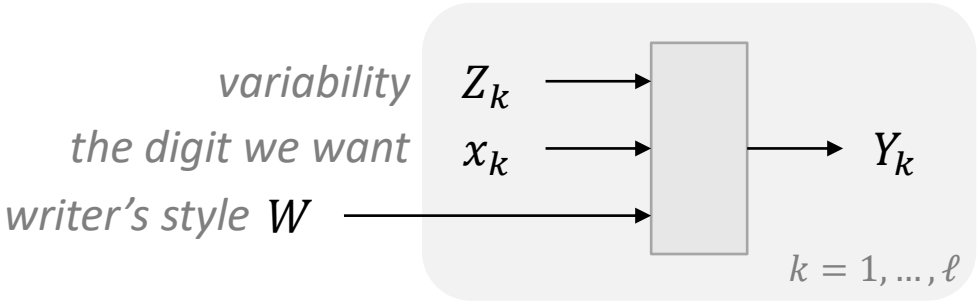




The dream: a handwriting generator that generates new characters in your style, as you type at the keyboard.

Why is our earlier CVAE handwriting generator not a great solution?

Grouped conditional generation



	$x = 0$	$x = 1$	$x = 2$	$x = 3$
writer $i = 1$				
writer $i = 2$				
writer $i = 3$				

In fact, MNIST is actually derived from a dataset of this form.
QMNIST, recovered by Yadav and Bottou (2019), restores the writer ids.

Our generative probability model for the data dictates the ‘wiring’ of our network

1. Propose a generative probability model that suits the dataset
2. Write down a KL-minimization problem
3. Deduce the wiring: which inputs need to go to which submodules
4. Train it!

TODAY ...
If we write our fitting problem with importance sampling,
then the wiring that we get out is an autoencoder.