



UPPSALA
UNIVERSITET

Machine learning – Block 7

Måns Magnusson
Department of Statistics, Uppsala University

Autumn 2022

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



UPPSALA
UNIVERSITET

Assignment 6: Evaluation

- **Practicalities**

- Introduction

- Autoencoders

- The Variational Autoencoder

- The probabilistic decoder
- The encoder
- Training a variational autoencoder

● ?

- Probabilistic Topic Models

- Latent Dirichlet Allocation
- Estimating the LDA model



UPPSALA
UNIVERSITET

This week's lectures

- **Practicalities**

- Introduction

- Autoencoders

- The Variational
Autoencoder

- The probabilistic decoder
- The encoder
- Training a variational
autoencoder

- Probabilistic Topic
Models

- Latent Dirichlet Allocation
- Estimating the LDA model

- Variational autoencoders
- Probabilistic Topic Models



UPPSALA
UNIVERSITET

- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Section 2

Introduction



Why variational autoencoders and topic models?

- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Popular approaches in **industry and academia**
- **Probabilistic** methods for unsupervised learning



Why variational autoencoders and topic models?

- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Popular approaches in **industry and academia**
- **Probabilistic** methods for unsupervised learning
- **Aim** of this lecture:
 - Describe the models
 - How to estimate these models
 - Explain what they are used for



UPPSALA
UNIVERSITET

Use Cases

- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of **images**
- Topic models: Unsupervised modeling of **documents**



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of images
- Topic models: Unsupervised modeling of documents
- Used for:
 - Identify "closeness" in high-dimensional data



- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of **images**
- Topic models: Unsupervised modeling of **documents**
- Used for:
 - Identify "**closeness**" in high-dimensional data
 - **Visualize/analyze** data



- Practicalities
 - **Introduction**
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of **images**
 - Topic models: Unsupervised modeling of **documents**
 - Used for:
 - Identify "**closeness**" in high-dimensional data
 - **Visualize/analyze** data
 - **Compression**



- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of **images**
- Topic models: Unsupervised modeling of **documents**
- Used for:
 - Identify "**closeness**" in high-dimensional data
 - **Visualize/analyze** data
 - **Compression**
 - **Feature construction**



- Practicalities
 - **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Variational autoencoders: Unsupervised modeling of **images**
- Topic models: Unsupervised modeling of **documents**
- Used for:
 - Identify "**closeness**" in high-dimensional data
 - **Visualize/analyze** data
 - **Compression**
 - **Feature construction**
 - Analyze underlying **patterns**



- Practicalities
- **Introduction**
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Use Cases: Examples

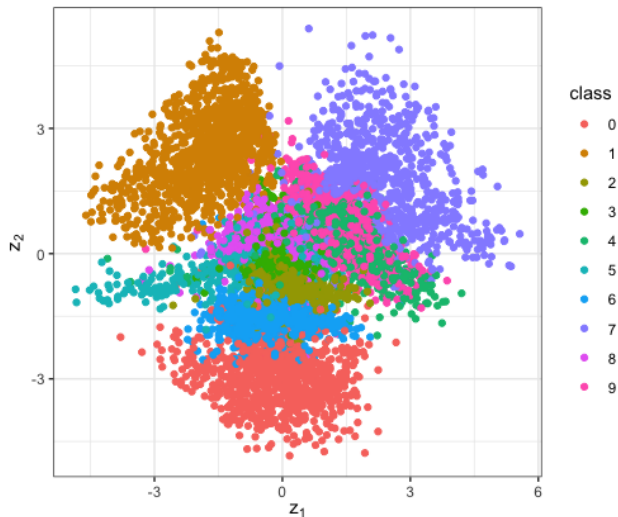


Figure: The latent state of MNIST using an Variational Autoencoder



UPPSALA
UNIVERSITET

- Practicalities
- Introduction
- **Autoencoders**
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Section 3

Autoencoders



Autoencoder

- An autoencoder is a neural network (e.g. feed-forward) that take an input x and predict (the same) x (r).

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Traing a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Autoencoder

- An autoencoder is a neural network (e.g. feed-forward) that take an input x and predict (the same) x (r).
- Three parts:
 - encoder $f(x)$ (or $e(x)$)
 - code h
 - decoder $g(h)$ (or $d(z)$)

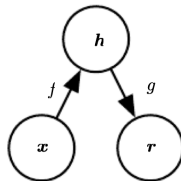


Figure: A Neural Autoencoder (Goodfellow et al, 2018)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Autoencoder

- An autoencoder is a neural network (e.g. feed-forward) that take an input x and predict (the same) x (r).
- Three parts:
 - encoder $f(x)$ (or $e(x)$)
 - code h
 - decoder $g(h)$ (or $d(z)$)

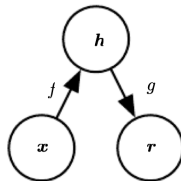


Figure: A Neural Autoencoder (Goodfellow et al, 2018)

- Loss function (reconstruction error):

$$L(\phi, \theta) = (x - d_{\phi}(e_{\theta}(x)))^2$$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Undercomplete Autoencoder

- More interesting: an **undercomplete** autoencoder:
Dimension of code is **lower** than that of x

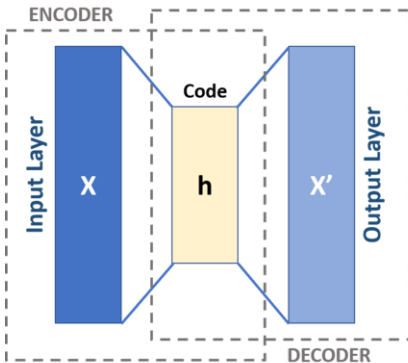


Figure: A Neural Autoencoder (Wikipedia)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

PCA and autoencoders

- A linear autoencoder: $e_\theta(x) = W_\phi$, and $d_\theta(x) = W_\phi$
- We want to minimize the loss (ignoring b /the mean):

$$L(\phi, \theta) = \sum_{i=1}^N (x_i - W_\theta W_\phi x_i)^2$$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

PCA and autoencoders

- A linear autoencoder: $e_\theta(x) = W_\phi$, and $d_\theta(x) = W_\phi$
- We want to minimize the loss (ignoring b /the mean):

$$L(\phi, \theta) = \sum_{i=1}^N (x_i - W_\theta W_\phi x_i)^2$$

- Remember **PCA loss**:

$$L(P) = \sum_{i=1}^N (x_i - P_q P_q^T x_i)^2,$$

where P is an orthogonal matrix of rank q .



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

PCA and autoencoders

- A linear autoencoder: $e_{\theta}(x) = W_{\phi}$, and $d_{\theta}(x) = W_{\phi}$
- We want to minimize the loss (ignoring b /the mean):

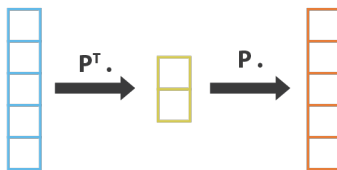
$$L(\phi, \theta) = \sum_{i=1}^N (x_i - W_{\theta} W_{\phi} x_i)^2$$

- Remember **PCA loss**:

$$L(P) = \sum_{i=1}^N (x_i - P_q P_q^T x_i)^2,$$

where P is an orthogonal matrix of rank q .

- **Hence:** PCA can be seen as an autoencoder





- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Deep Autoencoder: An autoencoder with multilayer neural networks as encoder and decoder
 - can be seen as a non-linear PCA
 - learn nonlinear representations



- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Deep Autoencoder: An autoencoder with multilayer neural networks as encoder and decoder
 - can be seen as a non-linear PCA
 - learn nonlinear representations
 - Problem: Deep autoencoders need to be regularized to not overfit the latent state



- Practicalities
 - Introduction
 - **Autoencoders**
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Problem: Autoencoders (as PCA) are not probabilistic models:
 - cannot generate data.
 - no notion of uncertainty



- Practicalities
- Introduction
- **Autoencoders**
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Problem: Autoencoders (as PCA) are not probabilistic models:
 - cannot generate data.
 - no notion of uncertainty
- We would like something like probabilistic PCA for (deep) autoencoders



- Remember the pPCA model (with z as latent variable):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i^T, \sigma I)$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

probabilistic PCA as an decoder

- Remember the pPCA model (with z as latent variable):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i^T, \sigma I)$$

- Now, swap the simple parameters with a neural network

$$x_i \sim N(\text{NeuralNetwork}_{\phi}(z_i), \sigma I)$$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

probabilistic PCA as an decoder

- Remember the pPCA model (with z as latent variable):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i^T, \sigma I)$$

- Now, swap the simple parameters with a neural network

$$x_i \sim N(\text{NeuralNetwork}_{\phi}(z_i), \sigma I)$$

- This is an example of a Deep Latent Variable model (a probabilistic decoder)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

probabilistic PCA as an decoder

- Remember the pPCA model (with z as latent variable):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i^T, \sigma I)$$

- Now, swap the simple parameters with a neural network

$$x_i \sim N(\text{NeuralNetwork}_{\phi}(z_i), \sigma I)$$

- This is an example of a Deep Latent Variable model (a probabilistic decoder)
- Another example is the Variational Autoencoder



UPPSALA
UNIVERSITET

- Practicalities
- Introduction
- Autoencoders
- **The Variational Autoencoder**
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Section 4

The Variational Autoencoder



UPPSALA
UNIVERSITET

The Variational Autoencoder (VAE)

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The variational autoencoder (VAE) is a **deep probabilistic autoencoder**
- Used for unsupervised learning of **images**



The Variational Autoencoder (VAE)

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The variational autoencoder (VAE) is a **deep probabilistic autoencoder**
- Used for unsupervised learning of **images**
- Consists of:
 1. The (probabilistic) encoder $q(z|\phi, x)$: **inference model**
 2. Sample z from encoded x : the **latent state**
 3. The (probabilistic) decoder $p(x|\theta, z)$: **observation model**



The Variational Autoencoder (VAE)

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The variational autoencoder (VAE) is a **deep probabilistic autoencoder**
- Used for unsupervised learning of **images**
- Consists of:
 1. The (probabilistic) encoder $q(z|\phi, x)$: **inference model**
 2. Sample z from encoded x : the **latent state**
 3. The (probabilistic) decoder $p(x|\theta, z)$: **observation model**
- Encoding the **latent state as a distribution** forces the space to be "reasonable"/reduces overfitting



The Variational Autoencoder (VAE)

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The variational autoencoder (VAE) is a **deep probabilistic autoencoder**
- Used for unsupervised learning of **images**
- Consists of:
 1. The (probabilistic) encoder $q(z|\phi, x)$: **inference model**
 2. Sample z from encoded x : the **latent state**
 3. The (probabilistic) decoder $p(x|\theta, z)$: **observation model**
- Encoding the **latent state as a distribution** forces the space to be "reasonable"/reduces overfitting
- VAEs get their name from **variational inference** (used in training)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Variational Autoencoder

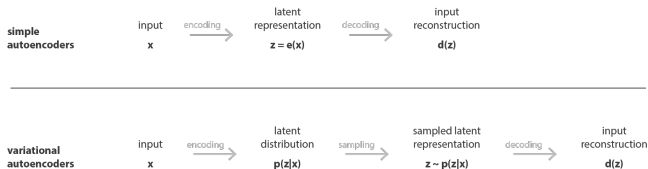


Figure: Autoencoder vs. the Variational Autoencoder (Rocca, 2019)



- Practicalities
- Introduction
- Autoencoders
- **The Variational Autoencoder**
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Variational Autoencoder

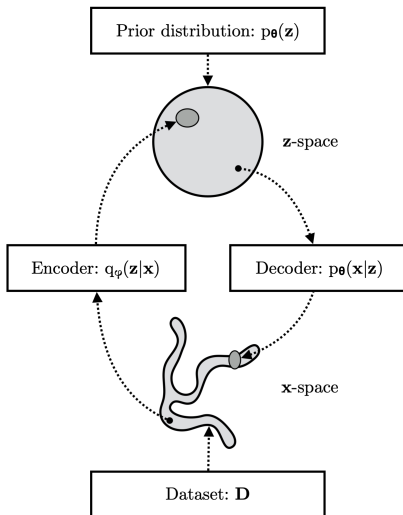


Figure: The Variational Autoencoder (Kingma and Welling, 2018, Fig. 2.1)



The probabilistic decoder

- The probabilistic decoder $p(x|\theta, z)$ (**observation model**)
- Usually a Normal distribution:

$$x_i \sim N(\text{NeuralNetwork}(z, \theta), cI)$$

- x_i for observation i depends non-linearly on the latent state z_i

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The probabilistic decoder

- The probabilistic decoder $p(x|\theta, z)$ (**observation model**)
- Usually a Normal distribution:

$$x_i \sim N(\text{NeuralNetwork}(z, \theta), cI)$$

- x_i for observation i depends non-linearly on the latent state z_i
- A probabilistic linear decoder: **pPCA**

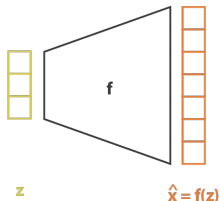


Figure: The Decoder (Rocca, 2019)



UPPSALA
UNIVERSITET

The probabilistic encoder

- The probabilistic encoder $q(z|x, \phi)$ (inference model)

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The probabilistic encoder $q(z|x, \phi)$ (inference model)
- We assume that $q_\phi(z|x)$ follows a specific distribution. Commonly:

$$z \sim N(\mu, \Sigma)$$

- A neural network learns the parameters μ and Σ

$$\mu = \text{NeuralNetwork}(x, \phi_\mu), \Sigma = \text{NeuralNetwork}(x, \phi_\Sigma),$$

where $\phi = (\phi_\mu, \phi_\Sigma)$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- The probabilistic encoder $q(z|x, \phi)$ (inference model)
- We assume that $q_\phi(z|x)$ follows a specific distribution. Commonly:

$$z \sim N(\mu, \Sigma)$$

- A neural network learns the parameters μ and Σ

$$\mu = \text{NeuralNetwork}(x, \phi_\mu), \Sigma = \text{NeuralNetwork}(x, \phi_\Sigma),$$

where $\phi = (\phi_\mu, \phi_\Sigma)$

- Common assumption: Σ is a diagonal matrix.



The probabilistic encoder

- The probabilistic encoder $q(z|x, \phi)$ (**inference model**)
- We assume that $q_\phi(z|x)$ follows a specific distribution. Commonly:

$$z \sim N(\mu, \Sigma)$$

- A neural network learns the parameters μ and Σ

$$\mu = \text{NeuralNetwork}(x, \phi_\mu), \Sigma = \text{NeuralNetwork}(x, \phi_\Sigma),$$

where $\phi = (\phi_\mu, \phi_\Sigma)$

- Common assumption: Σ is a diagonal matrix.
- **Result:** z_i for observation i depends non-linearly on x_i

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - **The encoder**
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - **The encoder**
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The probabilistic encoder

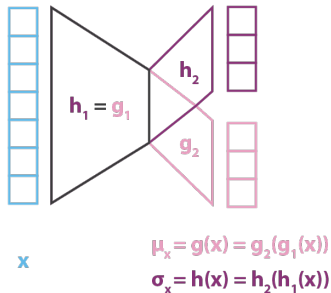


Figure: The Encoder (Rocca, 2019)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Variational Autoencoder

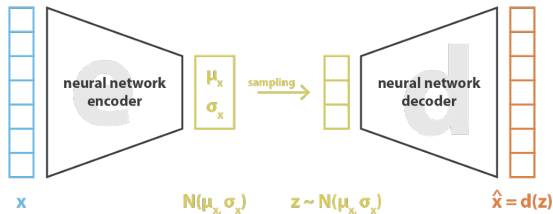


Figure: The Variational Autoencoder (Rocca, 2019)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - **The encoder**
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Variational Autoencoder

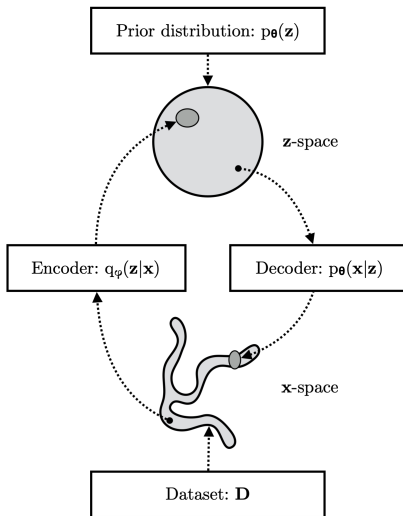


Figure: The Variational Autoencoder (Kingma and Welling, 2018, Fig. 2.1)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- **Goal:** estimating ϕ , θ (and z_i)
- The encoder and decoder are (usually) complicated (no close form solution)
- Need to estimate ϕ and θ using gradient ascent
- **Target:**
 - Maximize $\log p(x)$
(Explain the data as well as possible)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- **Goal:** estimating ϕ , θ (and z_i)
- The encoder and decoder are (usually) complicated (no close form solution)
- Need to estimate ϕ and θ using gradient ascent
- **Target:**
 - Maximize $\log p(x)$
(Explain the data as well as possible)
- **Optimization target:**
Maximize the variational lower bound or **evidence lower bound (ELBO)**



The marginal log-likelihood

$$\begin{aligned}\log p_{\theta}(x) &= \int q_{\phi}(z|x) \log p_{\theta}(x) dz \\ &= \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x)] \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{p_{\theta}(z|x)} \right) \right], \text{ using } p(z|x) = \frac{p(x, z)}{p(x)} \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right) \right] \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right) \right] + \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right) \right] \\ &= \underbrace{\mathcal{L}_{\theta, \phi}(x)}_{\text{ELBO}} + D_{KL}(q_{\phi}(z|x) || p_{\theta}(z|x))\end{aligned}$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



$$\begin{aligned}\log p_{\theta}(x) &= \int q_{\phi}(z|x) \log p_{\theta}(x) dz \\ &= \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x)] \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right) \right], \text{ using } p(z|x) = \frac{p(x, z)}{p_{\theta}(z|x)} \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right) \right] \\ &= \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right) \right] + \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right) \right] \\ &= \underbrace{\mathcal{L}_{\theta, \phi}(x)}_{\text{ELBO}} + D_{KL}(q_{\phi}(z|x) || p_{\theta}(z|x))\end{aligned}$$

$$\underbrace{\mathcal{L}_{\theta, \phi}(x)}_{\text{ELBO}} = \log p_{\theta}(x) - D_{KL}(q_{\phi}(z|x) || p_{\theta}(z|x))$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- **The Kullback-Leibler divergence:** a way of measuring the distance between probability distributions (although, not a metric)

$$D_{KL}(q_{\phi}(z|x)||p_{\theta}(z|x)) = \mathbb{E}_{q_{\phi}(z|x)} \left[\log \left(\frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right) \right]$$

$$D_{KL}(q_{\phi}(z|x)||p_{\theta}(z|x)) \geq 0$$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Optimization target: Maximize the ELBO

$$\mathcal{L}_{\theta, \phi}(x) = \log p_{\theta}(x) - D_{KL}(q_{\phi}(z|x) || p_{\theta}(z|x))$$

- ELBO is a lower bound for the marginal log-likelihood (similar to the EM algorithm)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Optimization target: Maximize the ELBO

$$\mathcal{L}_{\theta, \phi}(x) = \log p_{\theta}(x) - D_{KL}(q_{\phi}(z|x) || p_{\theta}(z|x))$$

- ELBO is a lower bound for the marginal log-likelihood (similar to the EM algorithm)
- Maximizing the ELBO will do two things:
 - Maximize the marginal log-likelihood $\log p_{\theta}(x)$:
Better generative model/decoder
 - Minimize the KL-divergence between $q_{\phi}(z|x)$ and $p_{\theta}(z|x)$:
Better approximation of the latent space/encoder



- Stochastic Gradient *Ascent* to maximize:

$$\begin{aligned}\mathcal{L}_{\theta,\phi}(x) &= \sum_i^N \mathcal{L}_{\theta,\phi}(x_i) \\ &= \sum_i^N \mathbb{E}_{q_{\phi}(z_i|x_i)} [\log(p_{\theta}(x_i, z_i)) - \log(q_{\phi}(z_i|x_i))]\end{aligned}$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Stochastic Gradient *Ascent* to maximize:

$$\begin{aligned}\mathcal{L}_{\theta,\phi}(x) &= \sum_i^N \mathcal{L}_{\theta,\phi}(x_i) \\ &= \sum_i^N \mathbb{E}_{q_{\phi}(z_i|x_i)} [\log(p_{\theta}(x_i, z_i)) - \log(q_{\phi}(z_i|x_i))]\end{aligned}$$

- Two problems:
 1. How do we compute the expectation?
Solution: **Monte Carlo Approximation**

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - **Training a variational autoencoder**
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Stochastic Gradient *Ascent* to maximize:

$$\begin{aligned}\mathcal{L}_{\theta, \phi}(x) &= \sum_i^N \mathcal{L}_{\theta, \phi}(x_i) \\ &= \sum_i^N \mathbb{E}_{q_{\phi}(z_i|x_i)} [\log(p_{\theta}(x_i, z_i)) - \log(q_{\phi}(z_i|x_i))]\end{aligned}$$

- Two problems:
 1. How do we compute the expectation?
Solution: **Monte Carlo Approximation**
 2. How compute the gradient wrt ϕ ?
Solution: **Change of variables**: $z = g(\epsilon, \phi, x)$
This is called **the reparametrization trick**

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - **Training a variational autoencoder**
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Using the reparametrization trick and Monte Carlo approximation, we get:

$$\begin{aligned}\mathcal{L}_{\theta, \phi}(x) &= \mathbb{E}_{q_{\phi}(z|x)} [\log(p_{\theta}(x, z)) - \log(q_{\phi}(z|x))] \\ &= \mathbb{E}_{p(\epsilon)} [\log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))] \\ &\approx \log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))\end{aligned}$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Using the reparametrization trick and Monte Carlo approximation, we get:

$$\begin{aligned}\mathcal{L}_{\theta, \phi}(x) &= \mathbb{E}_{q_{\phi}(z|x)} [\log(p_{\theta}(x, z)) - \log(q_{\phi}(z|x))] \\ &= \mathbb{E}_{p(\epsilon)} [\log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))] \\ &\approx \log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))\end{aligned}$$

- A common approach: do the MC approximation with only one sample per datapoint x_i .



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Optimizing the ELBO

- Using the reparametrization trick and Monte Carlo approximation, we get:

$$\begin{aligned}\mathcal{L}_{\theta, \phi}(x) &= \mathbb{E}_{q_{\phi}(z|x)} [\log(p_{\theta}(x, z)) - \log(q_{\phi}(z|x))] \\ &= \mathbb{E}_{p(\epsilon)} [\log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))] \\ &\approx \log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))\end{aligned}$$

- A common approach: do the MC approximation with only **one sample per datapoint x_i** .
- We approximate both $\mathcal{L}_{\theta, \phi}(x)$ and $\nabla \mathcal{L}_{\theta, \phi}(x)$



- Using the reparametrization trick and Monte Carlo approximation, we get:

$$\begin{aligned}\mathcal{L}_{\theta,\phi}(x) &= \mathbb{E}_{q_{\phi}(z|x)} [\log(p_{\theta}(x, z)) - \log(q_{\phi}(z|x))] \\ &= \mathbb{E}_{p(\epsilon)} [\log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))] \\ &\approx \log(p_{\theta}(x, g(\epsilon, \phi, x))) - \log(q_{\phi}(g(\epsilon, \phi, x)|x))\end{aligned}$$

- A common approach: do the MC approximation with only **one sample per datapoint x_i** .
- We approximate both $\mathcal{L}_{\theta,\phi}(x)$ and $\nabla \mathcal{L}_{\theta,\phi}(x)$
- Sometimes called a **doubly stochastic** algorithm.

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - **Training a variational autoencoder**
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Autoencoding Variational Bayes Algorithm

Algorithm 1: Stochastic optimization of the ELBO. Since noise originates from both the minibatch sampling and sampling of $p(\epsilon)$, this is a doubly stochastic optimization procedure. We also refer to this procedure as the *Auto-Encoding Variational Bayes* (AEVB) algorithm.

Data:

\mathcal{D} : Dataset

$q_\phi(\mathbf{z}|\mathbf{x})$: Inference model

$p_\theta(\mathbf{x}, \mathbf{z})$: Generative model

Result:

θ, ϕ : Learned parameters

$(\theta, \phi) \leftarrow$ Initialize parameters

while *SGD not converged* **do**

$\mathcal{M} \sim \mathcal{D}$ (Random minibatch of data)

$\epsilon \sim p(\epsilon)$ (Random noise for every datapoint in \mathcal{M})

 Compute $\tilde{\mathcal{L}}_{\theta, \phi}(\mathcal{M}, \epsilon)$ and its gradients $\nabla_{\theta, \phi} \tilde{\mathcal{L}}_{\theta, \phi}(\mathcal{M}, \epsilon)$

 Update θ and ϕ using SGD optimizer

end

Figure: The Autoencoding Variational Bayes Algorithm (Kingma and Welling, 2018, Algo. 1)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Autoencoding Variational Bayes Algorithm

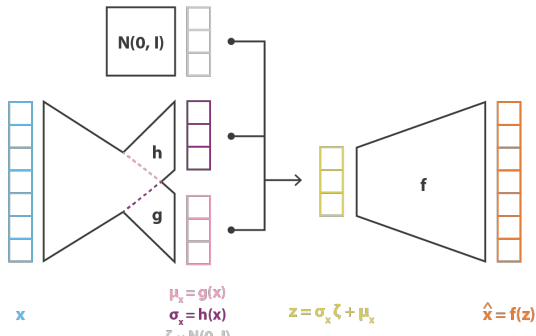


Figure: The Autoencoding Variational Bayes Algorithm (Rocca, 2019)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Summary

- Benefits of VAE:
 - Get a more interpretable latent state
 - We can estimate uncertainty
 - We can inject knowledge in our latent state



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Summary

- Benefits of VAE:
 - Get a more interpretable latent state
 - We can estimate uncertainty
 - We can inject knowledge in our latent state
- Problems:
 - The blurry image problem



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Summary

- Benefits of VAE:
 - Get a more interpretable latent state
 - We can estimate uncertainty
 - We can inject knowledge in our latent state
- Problems:
 - The blurry image problem
- Still much ongoing research:



Figure: Examples of images generated with a deep hierarchical Variational Autoencoder (Vahdat and Kautz, 2020)



UPPSALA
UNIVERSITET

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Section 5

Probabilistic Topic Models



UPPSALA
UNIVERSITET

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Section 5

Probabilistic Topic Models



UPPSALA
UNIVERSITET

Probabilistic Topic Models

- Unsupervised method for **textual data**

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



UPPSALA
UNIVERSITET

Probabilistic Topic Models

- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Unsupervised method for **textual data**
 - Popular in industry and academia to **analyze large corpora**



- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Unsupervised method for **textual data**
 - Popular in industry and academia to **analyze large corpora**
 - The most common model: **Latent Dirichlet Allocation**
 - A **mixed membership** model (a mixture of multinomial mixtures model)



- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Unsupervised method for **textual data**
 - Popular in industry and academia to **analyze large corpora**
 - The most common model: **Latent Dirichlet Allocation**
 - A **mixed membership** model (a mixture of multinomial mixtures model)
 - Topic model builds on the the **distributional hypothesis**



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Unsupervised method for **textual data**
- Popular in industry and academia to **analyze large corpora**
- The most common model: **Latent Dirichlet Allocation**
- A **mixed membership** model (a mixture of multinomial mixtures model)
- Topic model builds on the **distributional hypothesis**
- Use cases:
 - Create features for supervised models



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Unsupervised method for **textual data**
- Popular in industry and academia to **analyze large corpora**
- The most common model: **Latent Dirichlet Allocation**
- A **mixed membership** model (a mixture of multinomial mixtures model)
- Topic model builds on the the **distributional hypothesis**
- Use cases:
 - Create features for supervised models
 - Integrated in neural networks for model efficient learning



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Unsupervised method for **textual data**
- Popular in industry and academia to **analyze large corpora**
- The most common model: **Latent Dirichlet Allocation**
- A **mixed membership** model (a mixture of multinomial mixtures model)
- Topic model builds on the the **distributional hypothesis**
- Use cases:
 - Create features for supervised models
 - Integrated in neural networks for model efficient learning
 - Visualize document collections



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Unsupervised method for **textual data**
- Popular in industry and academia to **analyze large corpora**
- The most common model: **Latent Dirichlet Allocation**
- A **mixed membership** model (a mixture of multinomial mixtures model)
- Topic model builds on the **distributional hypothesis**
- Use cases:
 - Create features for supervised models
 - Integrated in neural networks for model efficient learning
 - Visualize document collections
 - Analyzing large corpora using statistical methods



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Unsupervised method for **textual data**
- Popular in industry and academia to **analyze large corpora**
- The most common model: **Latent Dirichlet Allocation**
- A **mixed membership** model (a mixture of multinomial mixtures model)
- Topic model builds on the the **distributional hypothesis**
- Use cases:
 - Create features for supervised models
 - Integrated in neural networks for model efficient learning
 - Visualize document collections
 - Analyzing large corpora using statistical methods
- Example: **All ears** media monitoring of speech data



The Dirichlet Distribution

- Probability distribution over the simplex with K categories:

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K x_i^{\alpha_i-1}$$

where

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)},$$

and where

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Dirichlet Distribution

- Probability distribution over the simplex with K categories:

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K x_i^{\alpha_i-1}$$

where

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)},$$

and where

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$$

- The probability distribution has the support on the simplex, that is

$$\sum_{i=1}^K x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in [1, K]$$



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Dirichlet Distribution

- Probability distribution over the simplex with K categories:

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K x_i^{\alpha_i-1}$$

where

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)},$$

and where

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$$

- The probability distribution has the support on the simplex, that is

$$\sum_{i=1}^K x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in [1, K]$$

- The parameters $\boldsymbol{\alpha}$ can be seen as **pseudo-counts**



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The Dirichlet Distribution

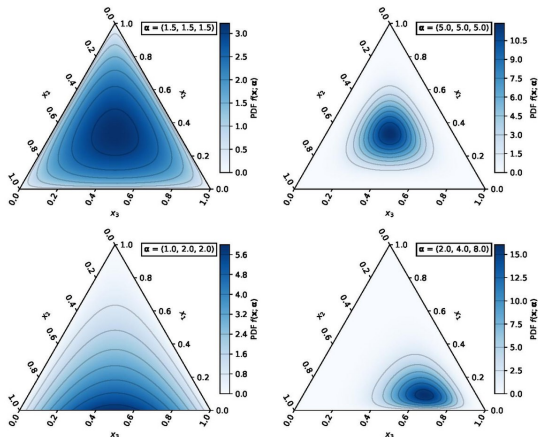


Figure: The Dirichlet Distribution (Wikipedia)



UPPSALA
UNIVERSITET

The distributional hypothesis

- Harris (1954) and Firths (1957):
“Word is characterized by the company it keeps”

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



The distributional hypothesis

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Harris (1954) and Firths (1957):
“Word is characterized by the company it keeps”
- Semantics (broadly defined) is captured by **context**



The distributional hypothesis

- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Harris (1954) and Firths (1957):
“Word is characterized by the company it keeps”
 - Semantics (broadly defined) is captured by **context**
 - Rough definition: **word windows** of different sizes



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

The distributional hypothesis

- Harris (1954) and Firths (1957):
“Word is characterized by the company it keeps”
- Semantics (broadly defined) is captured by **context**
- Rough definition: **word windows** of different sizes
- Different window sizes, different **semantic** content:
 - Word embeddings (context: word windows)
 - Topic models (context: documents)

Example

1. “A friend in need is a friend indeed.”
2. “She is my friend indeed.”



Latent Dirichlet Allocation

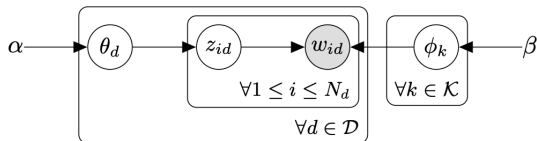


Figure: The Latent Dirichlet Allocation Model

where ϕ_k is the k th row in Φ (of dimension $K \times V$) and θ_d is the d th row in Θ (of dimension $D \times K$).



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Relies on the **bag-of-word** assumption

1. For each component k to K :

1.1 $\phi_k \sim \text{Dirichlet}(\beta)$

2. For each document d :

2.1 $\theta_d \sim \text{Dirichlet}(\alpha)$

2.2 For each token i :

2.2.1 $z_{id} \sim \text{Categorical}(\theta_d)$

2.2.2 $w_{id} \sim \text{Categorical}(\phi_{z_{id}})$



Example of parameters \mathbf{z} , Θ and Φ

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Example of parameters \mathbf{z} , Θ and Φ

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

| | | | | | | | |
|----------|---------|-------|-------|--------|--------|-------|-------|
| | | boat | shore | soccer | Zlatan | bank | money |
| $\Phi =$ | Topic 1 | 0.35 | 0.35 | 0.05 | 0.05 | 0.15 | 0.05 |
| | Topic 2 | 0.025 | 0.025 | 0.45 | 0.45 | 0.025 | 0.025 |
| | Topic 3 | 0.025 | 0.025 | 0.025 | 0.025 | 0.45 | 0.45 |



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Example of parameters \mathbf{z} , Θ and Φ

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

| | | | | | | | |
|----------|---------|-------|-------|--------|--------|-------|-------|
| | | boat | shore | soccer | Zlatan | bank | money |
| $\Phi =$ | Topic 1 | 0.35 | 0.35 | 0.05 | 0.05 | 0.15 | 0.05 |
| | Topic 2 | 0.025 | 0.025 | 0.45 | 0.45 | 0.025 | 0.025 |
| | Topic 3 | 0.025 | 0.025 | 0.025 | 0.025 | 0.45 | 0.45 |

| | | | | |
|------------|-------|---------|---------|---------|
| | | Topic 1 | Topic 2 | Topic 3 |
| $\Theta =$ | doc 1 | 0.96 | 0.02 | 0.02 |
| | doc 2 | 0.3 | 0.2 | 0.5 |
| | doc 3 | 0.05 | 0.35 | 0.6 |



UPPSALA UNIVERSITET

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

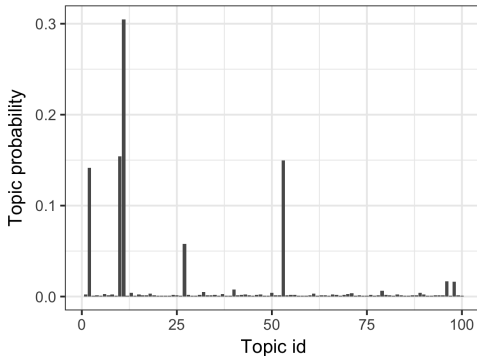
Closing arguments were heard yesterday in the Federal bankruptcy fraud trial of Stephen J. Sabbeth, whose legal problems have raised doubts about his ability to continue as leader of the Nassau County Democratic Party.

Mr. Sabbeth is charged with trying to conceal \$750,000 from his bank creditors by hiding the money in a secret account in his wife's maiden name, rather than use it to pay creditors when his lumber business went into bankruptcy 10 years ago.

— The New York Times 25th of February 1999



The estimated topic proportion ($\hat{\theta}_d$)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

| Topic | Top words (by ϕ_{kv}) |
|-------|--|
| 2 | party election voters campaign democratic |
| 10 | bank banks loans loan insurance savings |
| 11 | trial prison jury prosecutors convicted guilty |
| 53 | investigation inquiry documents investigators |

Table: The words with highest probability ($p(w|k)$) for topic 2, 10, 11 and 53.



The Latent Dirichlet Allocation Model

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

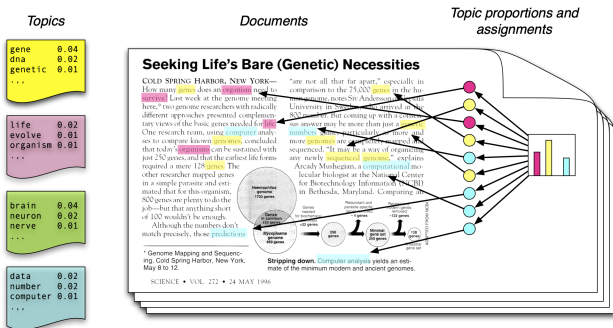


Figure: The Latent Dirichlet Allocation Model (Blei 2012, Fig. 1)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Common inference approaches
 1. Variational inference
 2. Markov Chain Monte Carlo (MCMC)



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Common inference approaches
 1. Variational inference
 2. Markov Chain Monte Carlo (MCMC)
- The Gibbs sampler is usually preferred
- Similar to (Stochastic) EM



The basic Gibbs sampler:

1. We want to estimate z, Φ, Θ :

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



The basic Gibbs sampler:

1. We want to estimate z, Φ, Θ :
2. Sample topic indicators (latent variable)

$$p(z = k | \Phi, \Theta) \propto \phi_{v,k} \theta_{k,d}$$

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Gibbs sampler for LDA

The basic Gibbs sampler:

1. We want to estimate z, Φ, Θ :
2. Sample topic indicators (latent variable)

$$p(z = k | \Phi, \Theta) \propto \phi_{v,k} \theta_{k,d}$$

3. Sample model parameters

$$\theta_d | \mathbf{z} \sim \text{Dir}(\mathbf{n}^{(d)} + \alpha)$$

$$\phi_k | \mathbf{z} \sim \text{Dir}(\mathbf{n}^{(v)} + \beta)$$

where $\mathbf{n}^{(d)}$ is the number of tokens by topic in document d and $\mathbf{n}^{(v)}$ is the number of tokens by topic for word type v .



Integrating out (collapsing) Θ and Φ

$$p(\mathbf{z}|\mathbf{w}) = \int \int p(\mathbf{z}, \Theta, \Phi | \mathbf{w}) \cdot p(\mathbf{z}, \Theta, \Phi) d\Phi d\Theta$$

will result in the following Gibbs sampler:

$$p(z_i = k | w_i, \mathbf{z}_{-i}) \propto \underbrace{\frac{n_k^{(v)} + \beta}{n_k^{(v)} + V\beta}}_{\text{type-topic } (\Phi)} \cdot \underbrace{(n_k^{(d)} + \alpha)}_{\text{topic-doc } (\Theta)}$$

where $n^{(v)}$ and $n^{(d)}$ are count matrices of size $D \times K$ and $K \times V$.

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



Example of $n^{(v)}$ and $n^{(d)}$

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



Example of $n^{(v)}$ and $n^{(d)}$

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

| | | | | | | |
|-------------|------|-------|--------|--------|------|-------|
| | boat | shore | soccer | Zlatan | bank | money |
| $n^{(v)} =$ | 2 | 2 | 0 | 0 | 1 | 0 |
| | 0 | 0 | 1 | 1 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 2 | 2 |

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Example of $n^{(v)}$ and $n^{(d)}$

| | | | | | |
|----------------|--------|-------|--------|-------|------|
| \mathbf{w}_1 | boat | shore | bank | | |
| \mathbf{z}_1 | 1 | 1 | 1 | | |
| \mathbf{w}_2 | Zlatan | boat | shore | money | bank |
| \mathbf{z}_2 | 2 | 1 | 1 | 3 | 3 |
| \mathbf{w}_3 | money | bank | soccer | money | |
| \mathbf{z}_3 | 3 | 3 | 2 | 3 | |

$$n^{(v)} = \begin{array}{ccccc} & \text{boat} & \text{shore} & \text{soccer} & \text{Zlatan} & \text{bank} & \text{money} \\ \begin{array}{c} 2 \\ 0 \\ 0 \end{array} & 2 & 2 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 1 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 2 & 2 \end{array}$$

$$n^{(d)} = \begin{bmatrix} 3 & 0 & 0 \\ 2 & 1 & 3 \\ 0 & 2 & 3 \end{bmatrix}$$



Topic Models as non-negative matrix factorization

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

$$\begin{bmatrix} n_{dv} \\ (D \times V) \end{bmatrix} \approx \begin{bmatrix} \Theta \\ (D \times K) \end{bmatrix} \times \begin{bmatrix} \Phi \\ (K \times V) \end{bmatrix}$$



UPPSALA
UNIVERSITET

Practicalities

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Setting K , α and β



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Setting K , α and β
- Reducing the vocabulary: stopwords, rare words, stemming



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Setting K , α and β
- Reducing the vocabulary: stopwords, rare words, stemming
- "Junk" topics



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Setting K , α and β
- Reducing the vocabulary: stopwords, rare words, stemming
- "Junk" topics
- We can analyze the topic indicators z directly



- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

Research Example: Swedish Immigration Discourse

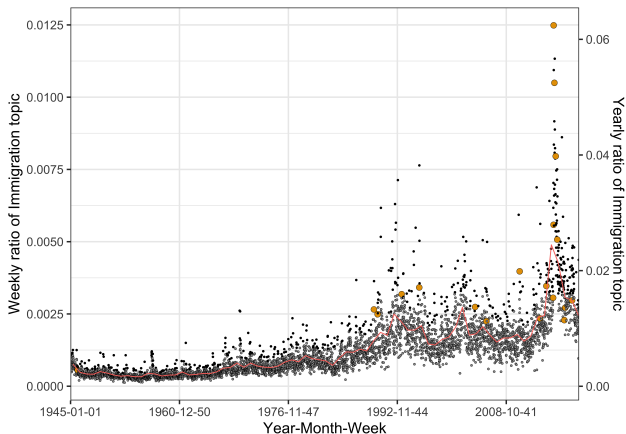


Figure: The Immigration topic in Swedish Newspapers (Hurtado Bodell et al, not in print)



UPPSALA
UNIVERSITET

Summary: Topic Models

- Practicalities
 - Introduction
 - Autoencoders
 - The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
 - Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model
- Topic models are unsupervised methods for textual data



UPPSALA
UNIVERSITET

Summary: Topic Models

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Topic models are unsupervised methods for textual data
- The Latent Dirichlet Allocation is a popular model



Summary: Topic Models

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Topic models are unsupervised methods for textual data
- The Latent Dirichlet Allocation is a popular model
- A mixed membership model (a mixture of multinomial mixtures model)



Summary: Topic Models

- Practicalities
- Introduction
- Autoencoders
- The Variational Autoencoder
 - The probabilistic decoder
 - The encoder
 - Training a variational autoencoder
- Probabilistic Topic Models
 - Latent Dirichlet Allocation
 - Estimating the LDA model

- Topic models are unsupervised methods for textual data
- The Latent Dirichlet Allocation is a popular model
- A mixed membership model (a mixture of multinomial mixtures model)
- Usually use Gibbs samplers for estimation