



UPPSALA  
UNIVERSITET

# Machine learning – Block 7

Måns Magnusson  
Department of Statistics, Uppsala University

Autumn 2025

- 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
- (Diffusion models)
- Large Language 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



- 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

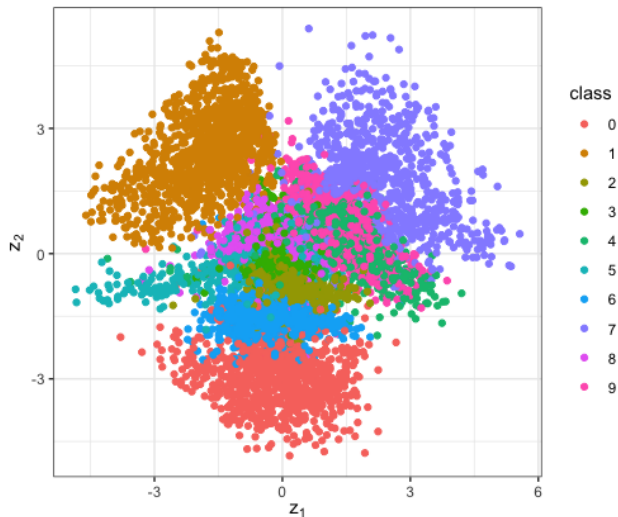


Figure: The latent state of MNIST using an Variational Autoencoder



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



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

- 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

# Autoencoder

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

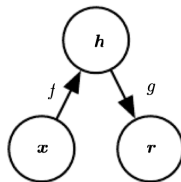


Figure: A Neural Autoencoder (Goodfellow et al, 2018)



# Autoencoder loss (Reconstruction error)

For a deterministic autoencoder:

$$h = e_{\phi}(x), \quad \hat{x} = d_{\theta}(h)$$

A common reconstruction loss is:

$$L(\theta, \phi) = \|x - \hat{x}\|_2^2 = \|x - d_{\theta}(e_{\phi}(x))\|_2^2$$

Minimizing squared error is equivalent (up to constants) to maximizing the log-likelihood of a Gaussian decoder with fixed variance:

$$x \mid h \sim \mathcal{N}(d_{\theta}(h), \sigma^2 I).$$

- 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 Undercomplete Autoencoder

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

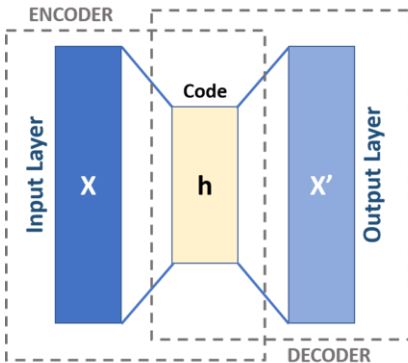


Figure: A Neural Autoencoder (Wikipedia)



# PCA and autoencoders

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

$$L(\theta, \phi) = \sum_{i=1}^N \|x_i - W_\theta W_\phi x_i\|_2^2$$

- 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

## PCA and autoencoders

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

$$L(\theta, \phi) = \sum_{i=1}^N \|x_i - W_\theta W_\phi x_i\|_2^2$$

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

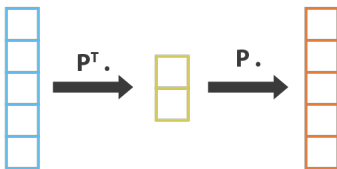
$$L(\theta, \phi) = \sum_{i=1}^N \|x_i - W_\theta W_\phi x_i\|_2^2$$

- 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 linear autoencoder (under squared loss and orthonormal constraints)





- 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):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i, \sigma^2 \mathbf{I})$$

- 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):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i, \sigma^2 I)$$

- Now, swap the simple parameters with a neural network

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



- 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):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i, \sigma^2 \mathbf{I})$$

- Now, swap the simple parameters with a neural network

$$x_i \sim N(\text{NeuralNetwork}_{\theta}(z_i), \sigma^2 \mathbf{I})$$

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



- 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):

$$x_i \sim N(\mathbf{b} + \mathbf{W}z_i, \sigma^2 \mathbf{I})$$

- Now, swap the simple parameters with a neural network

$$x_i \sim N(\text{NeuralNetwork}_{\theta}(z_i), \sigma^2 \mathbf{I})$$

- This is an example of a Deep Latent Variable model (DLVM)
- One example of a DLVM 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**



# 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**



# 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

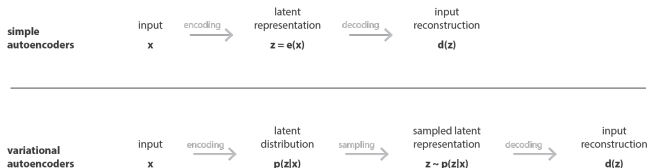
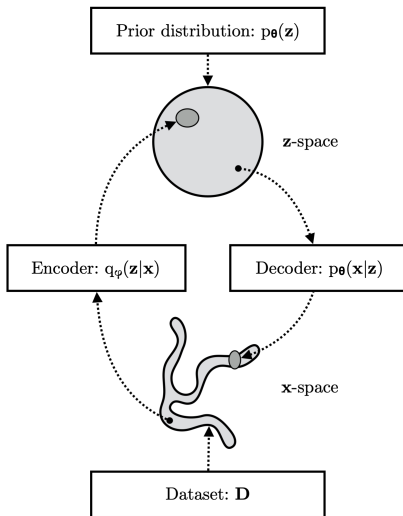


Figure: Autoencoder vs. the Variational Autoencoder (Rocca, 2019)



- 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**

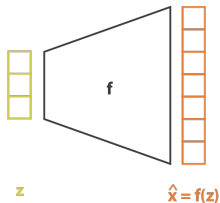


Figure: The Decoder (Rocca, 2019)



# 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

---

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

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



- 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

- 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  $\mu$  and  $\Sigma$

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

and

$$\log \sigma^2(x) = \text{NeuralNetwork}(x, \phi_\Sigma).$$



- 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

- 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  $\mu$  and  $\Sigma$

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

and

$$\log \sigma^2(x) = \text{NeuralNetwork}(x, \phi_\Sigma).$$

- **Result:**  $z_i$  for observation  $i$  depends non-linearly on  $x_i$



# The probabilistic encoder

- 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

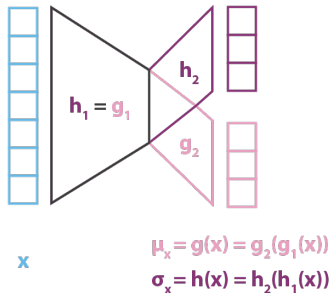


Figure: The Encoder (Rocca, 2019)



- 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

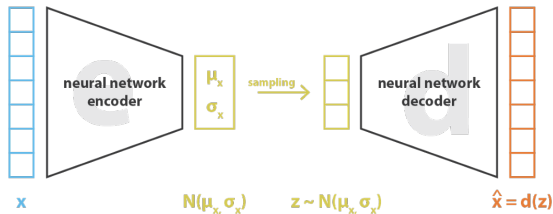
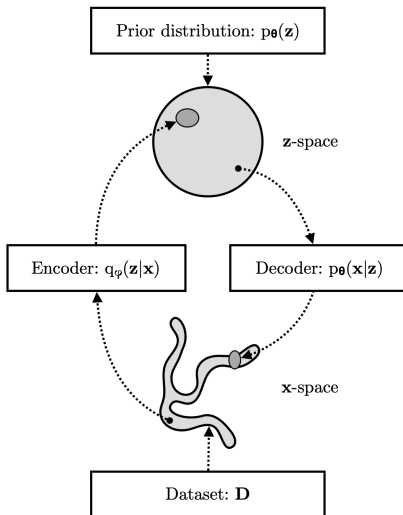


Figure: The Variational Autoencoder (Rocca, 2019)



- 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)



UPPSALA  
UNIVERSITET

# Training a 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

- **Goal:** estimating  $\phi$ ,  $\theta$  (and  $z_i$ )



- 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

- **Goal:** estimating  $\phi$ ,  $\theta$  (and  $z_i$ )
- The encoder and decoder are (usually) complicated (no close form solution)
- Estimate  $\phi$  and  $\theta$



- 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  $\phi$ ,  $\theta$  (and  $z_i$ )
  - The encoder and decoder are (usually) complicated (no close form solution)
  - Estimate  $\phi$  and  $\theta$
  - Target:
  - Maximize  $\log p(x)$  (i.e. 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

# The Kullback–Leibler (KL) Divergence

## What is it?

A measure of how well one probability distribution approximates another.

## Definition

For two distributions  $p$  and  $q$ :

$$D_{\text{KL}}(p \parallel q) = \mathbb{E}_p \left[ \log \frac{p(X)}{q(X)} \right]$$

## Key properties

- $D_{\text{KL}}(p \parallel q) \geq 0$
- $D_{\text{KL}}(p \parallel q) = 0$  iff  $p = q$
- Not symmetric:  $D_{\text{KL}}(p \parallel q) \neq D_{\text{KL}}(q \parallel p)$

## Interpretation

- Measures information lost when using  $q$  to approximate  $p$
- Direction matters: *which distribution is the truth*



# From the marginal likelihood to the ELBO

**Goal:** Compute or maximize the marginal log-likelihood

$$\log p_{\theta}(x)$$

**Step 1: Introduce the latent variable**

$$\log p_{\theta}(x) = \log \int p_{\theta}(x, z) dz$$

**Step 2: Introduce a variational distribution**

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

(This equality always holds as long as  $q_{\phi}(z | x) > 0$ .)

**Step 3: Apply Jensen's inequality**

(For a concave function  $f$ ,  $f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$ )

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

- 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 Evidence Lower Bound (ELBO)

**Definition:** The Evidence Lower Bound (ELBO) is

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

**Key relationship:**

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

(KL divergence is always non-negative.)

**Interpretation:**

- Maximizing the ELBO *increases*  $\log p_{\theta}(x)$
- Minimizes the gap between  $q_{\phi}(z | x)$  and  $p_{\theta}(z | x)$

**Important:**

$$\mathcal{L}_{\theta, \phi}(x) \leq \log p_{\theta}(x), \quad \text{with equality iff } q_{\phi}(z | x) = p_{\theta}(z | x)$$



# Optimizing the ELBO

---

## Objective

We want to maximize the ELBO with respect to encoder and decoder parameters:

$$\max_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x, z) - \log q_{\phi}(z | x)]$$

- 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



# Optimizing the ELBO

## Objective

We want to maximize the ELBO with respect to encoder and decoder parameters:

$$\max_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x, z) - \log q_{\phi}(z | x)]$$

## Problem

- The expectation is over  $q_{\phi}(z | x)$
- We need gradients with respect to  $\phi$  and  $\theta$
- Direct differentiation through sampling is not 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



- 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

## Objective

We want to maximize the ELBO with respect to encoder and decoder parameters:

$$\max_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x, z) - \log q_{\phi}(z | x)]$$

## Problem

- The expectation is over  $q_{\phi}(z | x)$
- We need gradients with respect to  $\phi$  and  $\theta$
- Direct differentiation through sampling is not possible

## Solution

- Approximate the expectation using Monte Carlo
- Rewrite the sampling step so gradients can flow (reparameterization 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

# The Reparameterization Trick

## Assumption

The variational posterior is Gaussian:

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

## Key idea

Instead of sampling  $z \sim q_{\phi}(z \mid x)$ , write:

$$z = g(\phi, x, \epsilon) = \mu_{\phi}(x) + \sigma_{\phi}(x) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

## Why this helps

- Randomness is now isolated in  $\epsilon$
- $z = g(\phi, x, \epsilon)$  is a differentiable function of  $\phi$
- Enables backpropagation through the sampling step



# Monte Carlo Optimization of the ELBO

## ELBO rewritten

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

## Monte Carlo approximation

Using one sample  $\epsilon^{(1)} \sim \mathcal{N}(0, 1)$ :

$$\mathcal{L}_{\theta, \phi}(x) \approx \log p_{\theta}(x, g(\phi, x, \epsilon^{(1)})) - \log q_{\phi}(g(\phi, x, \epsilon^{(1)}) | x)$$

## Optimization

- Use stochastic gradient ascent (or descent on  $-\text{ELBO}$ )
- Gradients flow through  $\mu_{\phi}(x)$  and  $\sigma_{\phi}(x)$  and we can compute  $\nabla \mathcal{L}_{\theta, \phi}(x)$
- One or few samples  $\epsilon$  per datapoint is usually sufficient

- 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:**

$\theta, \phi$ : 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  $\theta$  and  $\phi$  using SGD optimizer

**end**

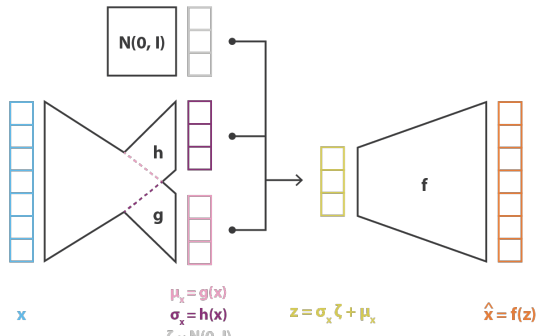
---

**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)



- Benefits of VAE:
  - Get a more **interpretable** latent state
  - We can estimate **uncertainty** (but its usually bad)
  - **Regularize** the latent state
  - 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



- 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** (but its usually bad)
  - **Regularize** the latent state
  - We can inject knowledge in our latent state
- Still much ongoing research



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



- 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

# VAE vs Diffusion Models

## Variational Autoencoders (VAEs)

- Latent variable model:  $z \sim p(z)$
- Trained by maximizing the ELBO
- Fast sampling (one forward pass)
- Compact, interpretable latent space
- Often produce blurrier samples

## Diffusion Models

- No explicit low-dimensional latent space
- Learn to reverse a noise process
- Trained via denoising objectives
- Very high-quality samples
- Slow sampling (many steps)

**Takeaway:** VAEs prioritize efficient inference and representation learning, while diffusion models prioritize sample quality.



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

---

- 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**



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**



# 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**
  - 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 **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:
  - **Explore** 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:
  - **Explore** 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 **distributional hypothesis**
- Use cases:
  - **Explore** document collections
  - **Analyzing large corpora** using statistical methods
- Example: **All ears** media monitoring of speech data



- Probability distribution over the simplex with  $K$  categories:

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

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

- 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:

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

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

- 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:

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

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

- 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** and

$$\mathbb{E}[x_i] = \frac{\alpha_i}{\sum_{j=1}^K \alpha_j}$$



- 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

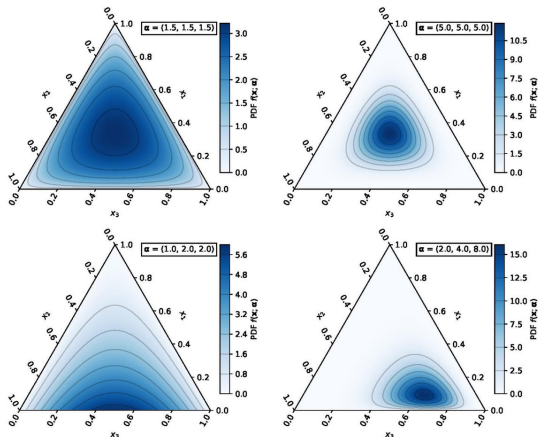


Figure: The Dirichlet Distribution (Wikipedia)



# 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 Firth (1957):  
“Word is characterized by the company it keeps”



# 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 Firth (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 Firth (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

---

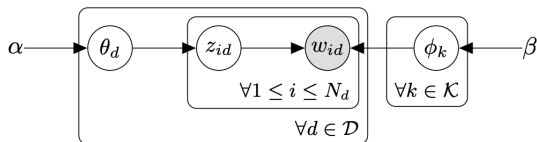
- Harris (1954) and Firth (1957):  
“Word is characterized by the company it keeps”
- Semantics (broadly defined) is captured by **context**
- Rough definition: **word windows** of different sizes
- Different window sizes, different **semantic** content:
  - Word embeddings (context: word windows)
  - 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  $\phi_k$  is the  $k$ th row in  $\Phi$  (of dimension  $K \times V$ ) and  $\theta_d$  is the  $d$ th row in  $\Theta$  (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  $w_{id} \sim \text{Categorical}(\phi_{z_{id}})$

- 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 $\mathbf{z}$ , $\Theta$ and $\Phi$

$\mathbf{w}_1$	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 $\mathbf{z}$ , $\Theta$ and $\Phi$

$\mathbf{w}_1$	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 $\mathbf{z}$ , $\Theta$ and $\Phi$

$\mathbf{w}_1$	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

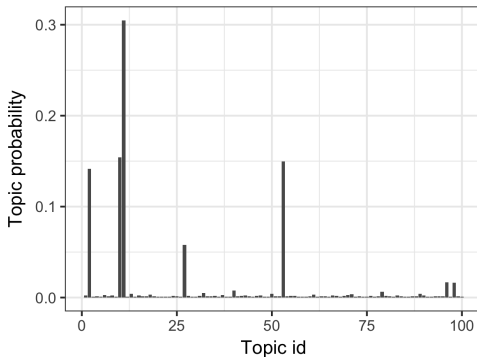
Closing arguments were heard yesterday in the Federal bankruptcy fraud trial of Stephen J. Sabbeth, whose legal problems have raised doubts about his ability to continue as leader of the Nassau County Democratic Party.

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

Topic	Top words (by $\phi_{kv}$ )
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

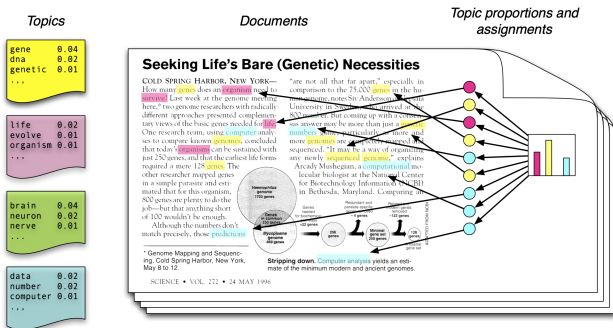


Figure: The Latent Dirichlet Allocation Model (Blei 2012, Fig. 1)



- 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



- 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  $z, \Phi, \Theta$ :

- 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  $z, \Phi, \Theta$ :
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



The basic Gibbs sampler:

1. We want to estimate  $z, \Phi, \Theta$ :
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  $v$ .

- 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

Integrating out (collapsing)  $\Theta$  and  $\Phi$

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

will result in the following Gibbs sampler:

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

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



## 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



## 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

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

- 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

# 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  $K$ ,  $\alpha$  and  $\beta$



- 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  $K$ ,  $\alpha$  and  $\beta$
- 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  $K$ ,  $\alpha$  and  $\beta$
  - 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

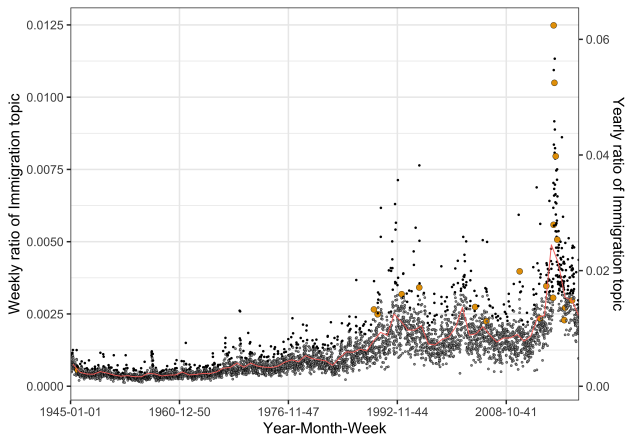
- Setting  $K$ ,  $\alpha$  and  $\beta$
- Reducing the vocabulary: stopwords, rare words, stemming
- "Junk" topics
- We can analyze the topic indicators  $z$  directly



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

# Research Example: Swedish Immigration Discourse



**Figure:** The Immigration topic in Swedish Newspapers (Hurtado Bodell et al, 2024)



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** models 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** models 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** models 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** models for textual data
  - The **Latent Dirichlet Allocation** is a popular model
  - A **mixed membership model** (a mixture of multinomial mixtures model)
  - Use Gibbs samplers for estimation