



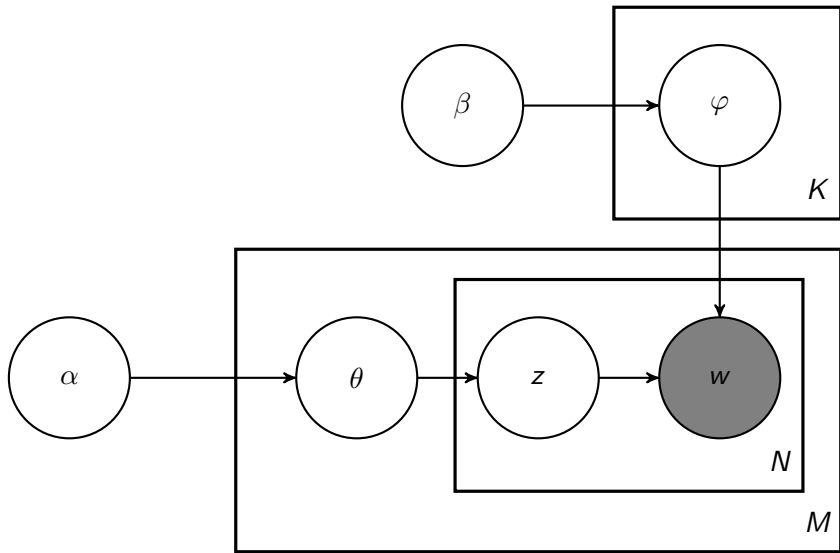
Parameter Estimation for the  
Latent Dirichlet Allocation  
SiKDD

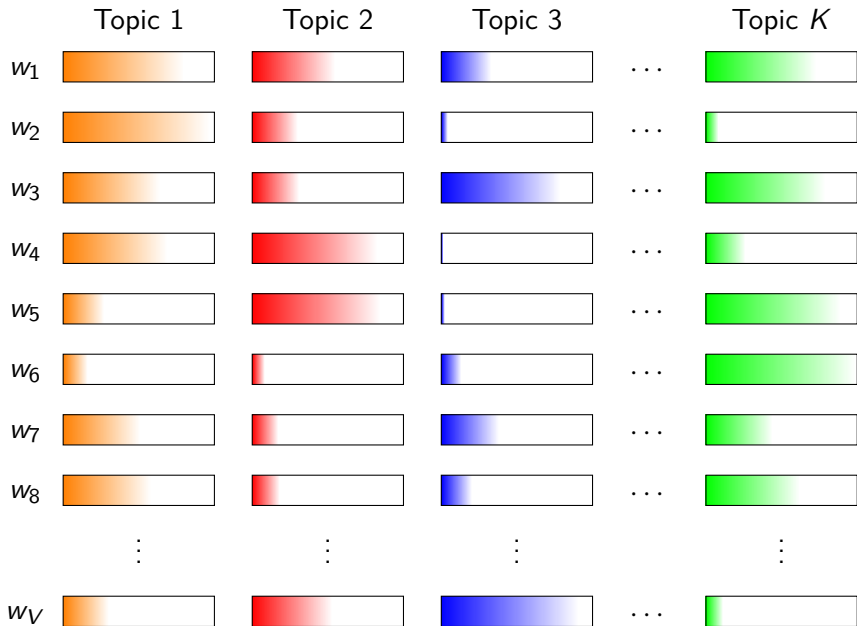
Jaka Špeh    Andrej Muhič    Jan Rupnik

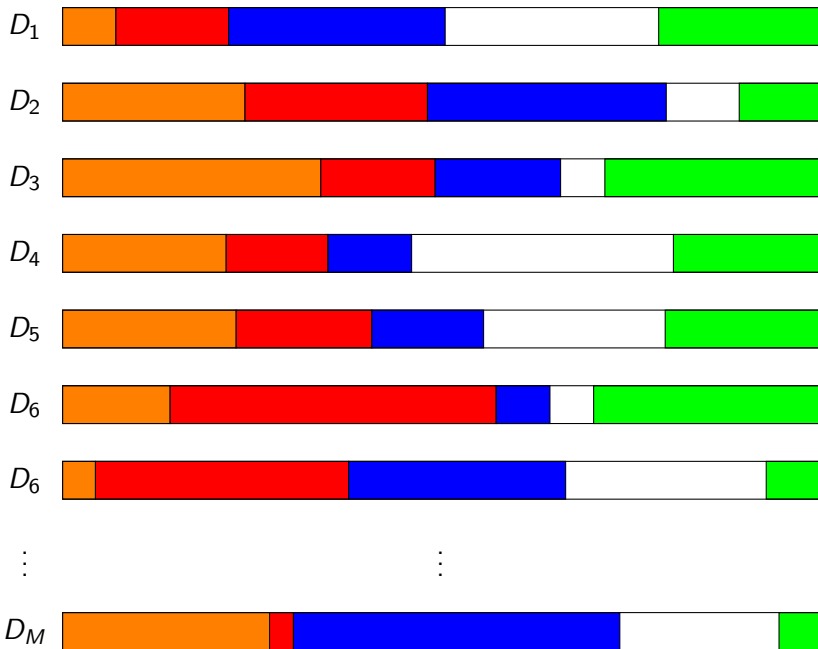
AI Laboratory  
Jozef Stefan Institute

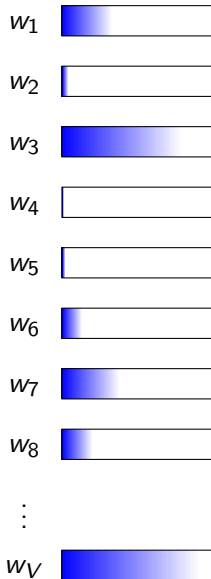
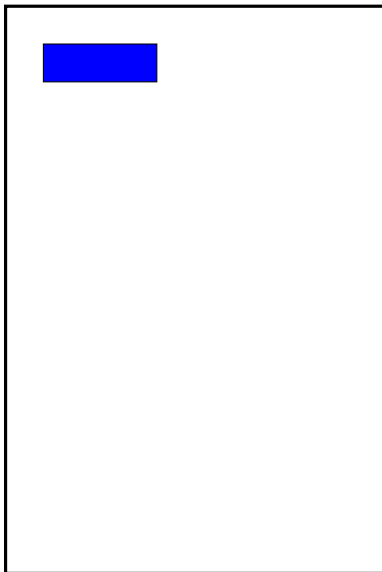
- LDA model.
- Algorithms.
- Comparison.

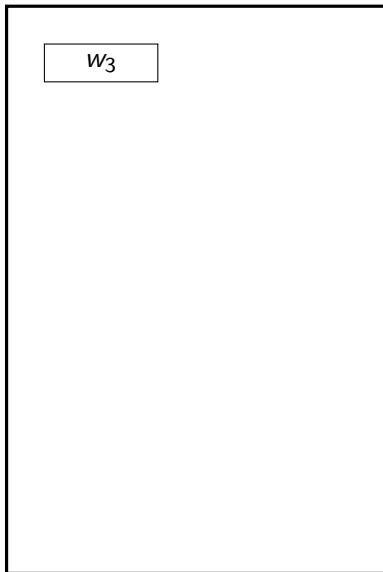

LDA model.



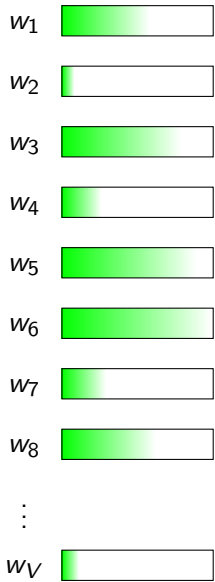
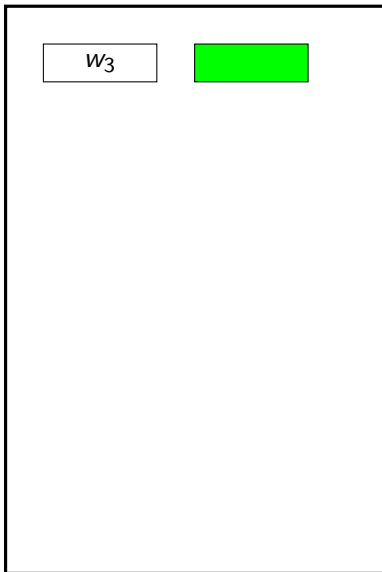


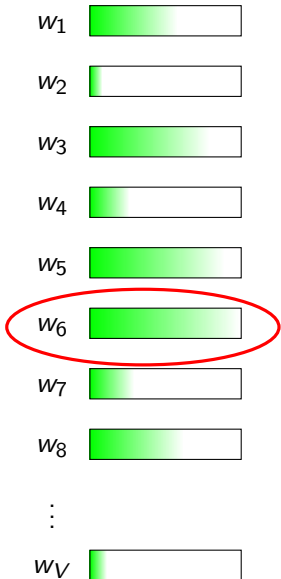
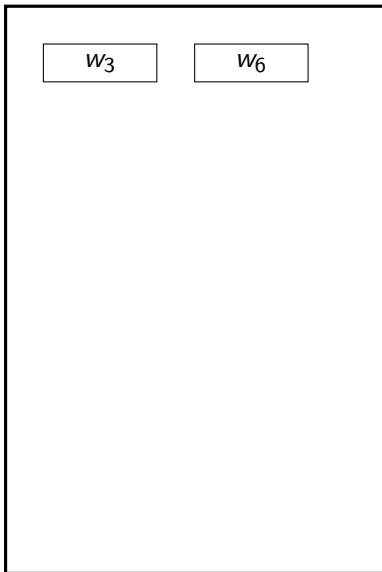


$D_1$ 

$D_1$  $w_1$   $w_2$   $w_3$   $w_4$   $w_5$   $w_6$   $w_7$   $w_8$   $\vdots$  $w_V$  





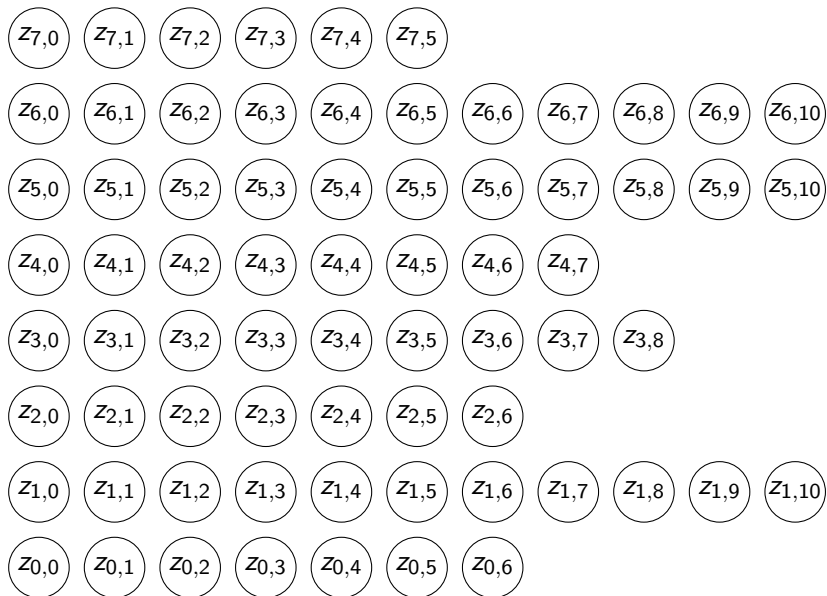


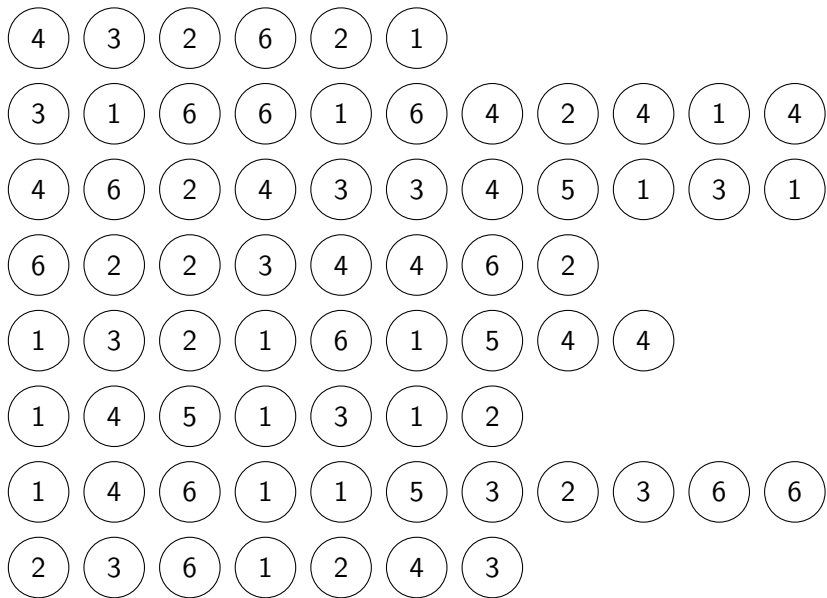
Algorithms.

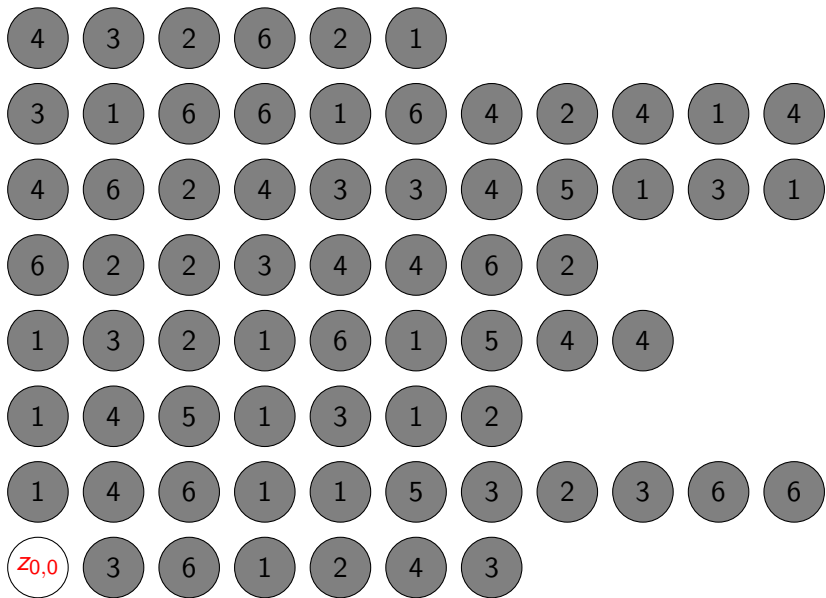
# Algorithms.

- Collapsed Gibbs Sampling.
- Variational Bayesian Inference.
- Online Variational Bayesian Inference.

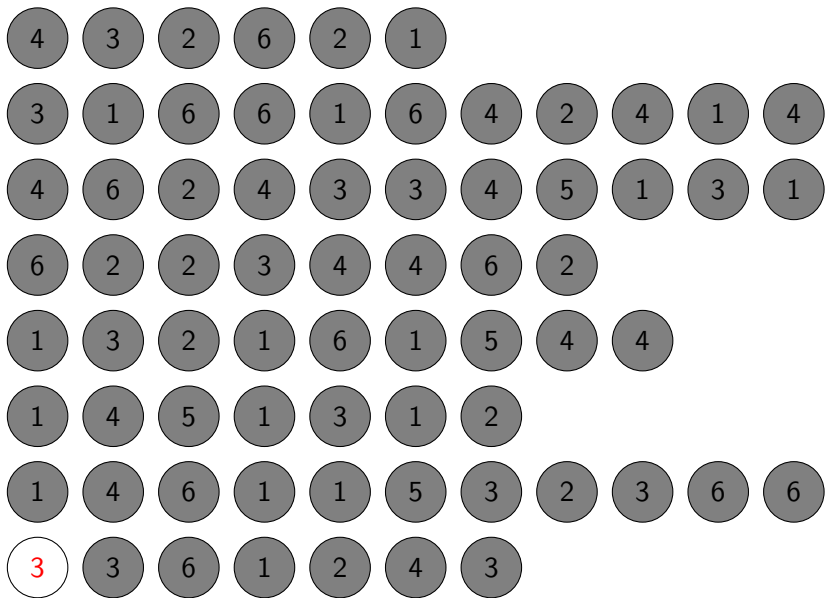
Collapsed Gibbs Sampling.

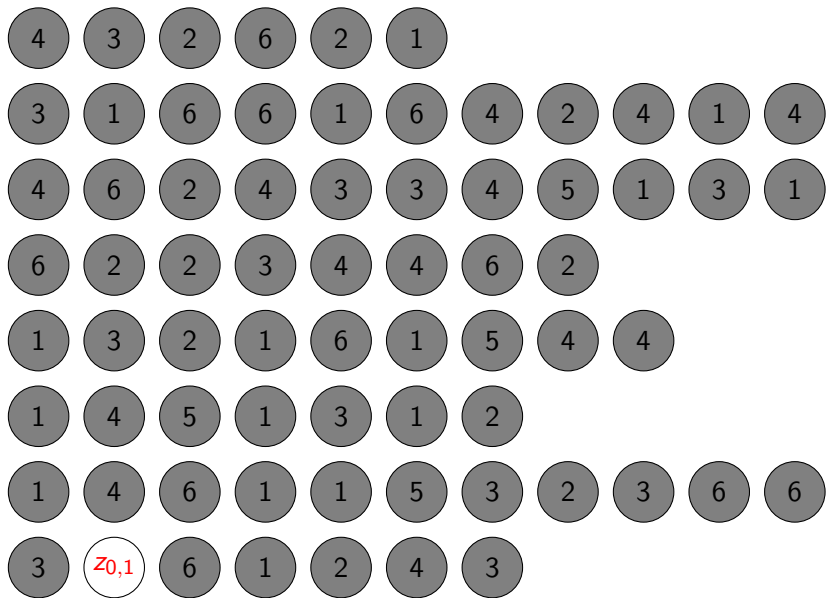


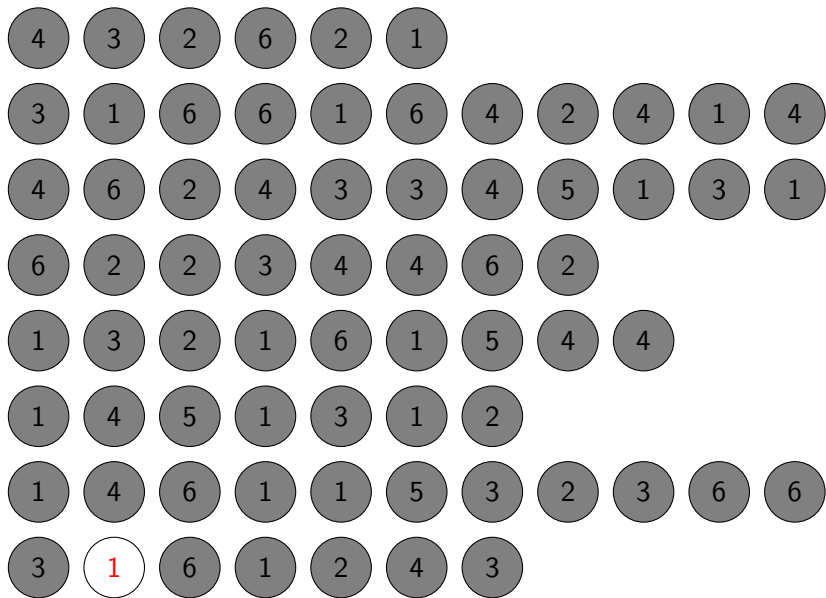


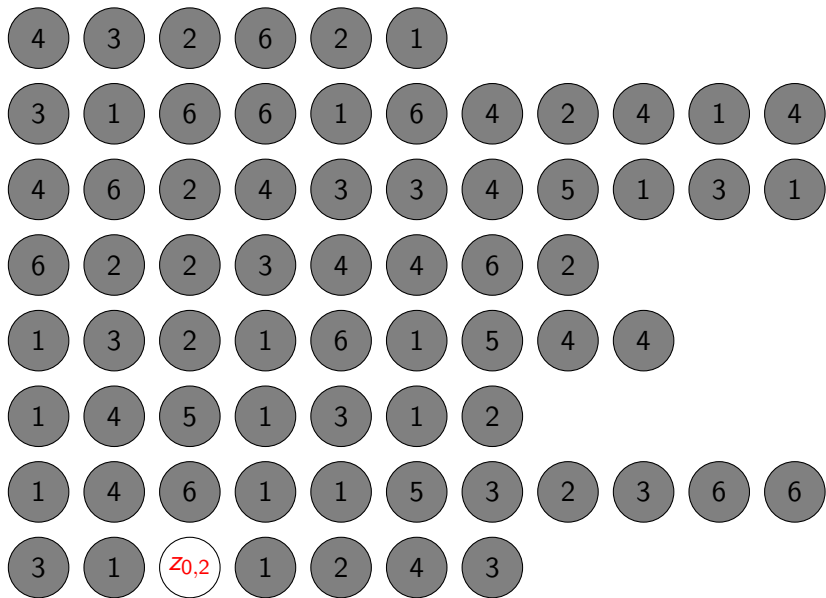






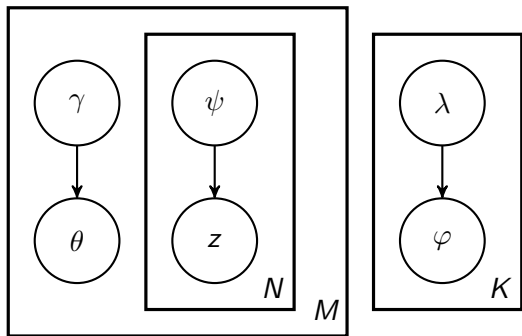
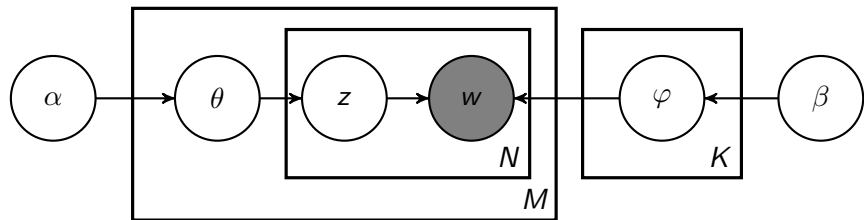






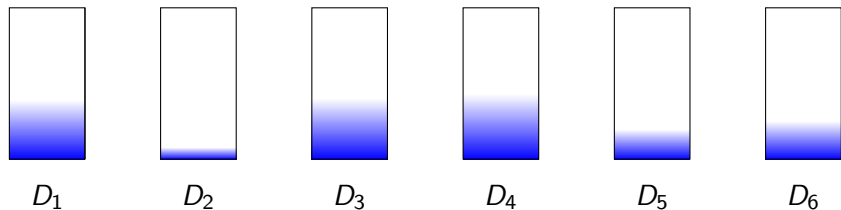
# Variational Bayesian Inference.

# Variational Bayesian Inference

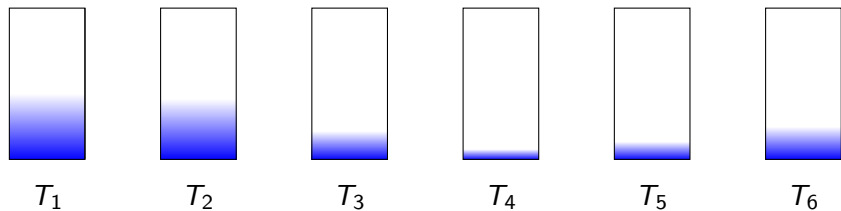


# VB: Initial State

Documents (ELBO contribution)

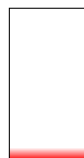


Topics (ELBO contribution)

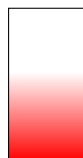


# VB: Step 1.

Documents



$D_1$



$D_2$



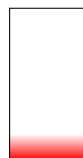
$D_3$



$D_4$



$D_5$

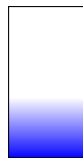


$D_6$

Topics



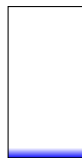
$T_1$



$T_2$



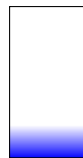
$T_3$



$T_4$



$T_5$



$T_6$



## VB: Step 2.

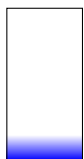
Documents



$D_1$



$D_2$



$D_3$



$D_4$

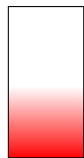


$D_5$



$D_6$

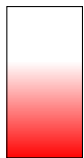
Topics



$T_1$



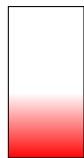
$T_2$



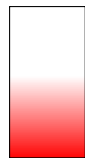
$T_3$



$T_4$



$T_5$



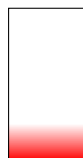
$T_6$

# VB: Step 3.

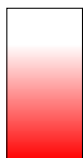
Documents



$D_1$



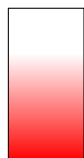
$D_2$



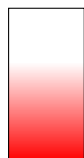
$D_3$



$D_4$



$D_5$

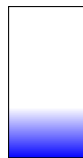


$D_6$

Topics



$T_1$



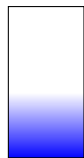
$T_2$



$T_3$



$T_4$



$T_5$



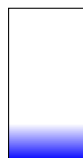
$T_6$

# VB: Step 4.

Documents



$D_1$



$D_2$



$D_3$



$D_4$

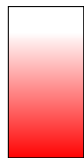


$D_5$



$D_6$

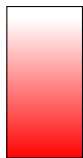
Topics



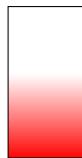
$T_1$



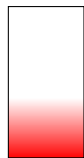
$T_2$



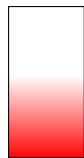
$T_3$



$T_4$



$T_5$

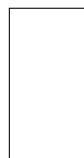


$T_6$

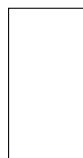
# Online Variational Bayesian Inference.

# Online VB: Initial Step

Documents



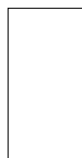
$D_1$



$D_2$



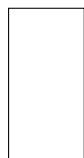
$D_3$



$D_4$

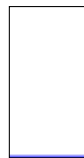


$D_5$

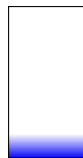


$D_6$

Topics



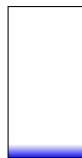
$T_1$



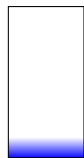
$T_2$



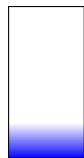
$T_3$



$T_4$



$T_5$



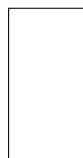
$T_6$

# Online VB: Step 1

Documents



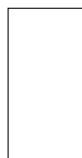
$D_1$



$D_2$



$D_3$



$D_4$

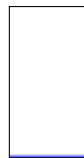


$D_5$

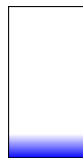


$D_6$

Topics



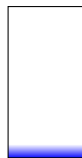
$T_1$



$T_2$



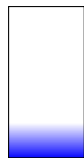
$T_3$



$T_4$



$T_5$



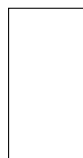
$T_6$

# Online VB: Step 2

Documents



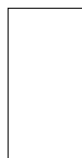
$D_1$



$D_2$



$D_3$



$D_4$



$D_5$

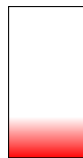


$D_6$

Topics



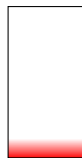
$T_1$



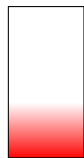
$T_2$



$T_3$



$T_4$



$T_5$



$T_6$

# Online VB: Step 3

Documents



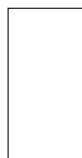
$D_1$



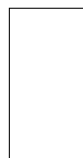
$D_2$



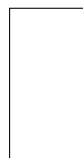
$D_3$



$D_4$

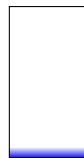


$D_5$

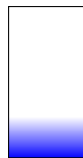


$D_6$

Topics



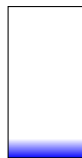
$T_1$



$T_2$



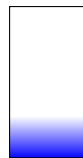
$T_3$



$T_4$



$T_5$



$T_6$



# Online VB: Step 4

Documents



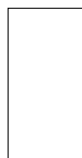
$D_1$



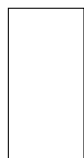
$D_2$



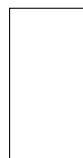
$D_3$



$D_4$



$D_5$

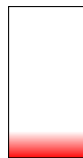


$D_6$

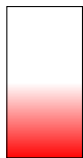
Topics



$T_1$



$T_2$



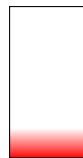
$T_3$



$T_4$



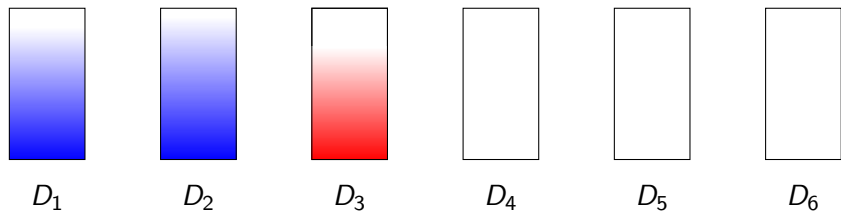
$T_5$



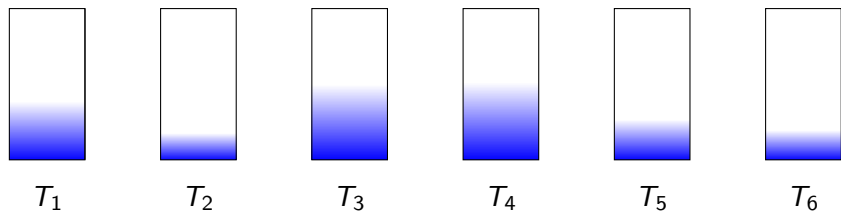
$T_6$

# Online VB: Step 5

Documents

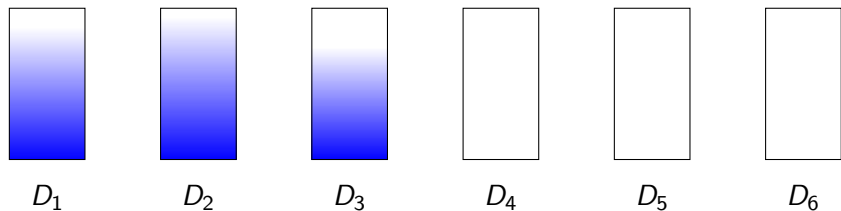


Topics

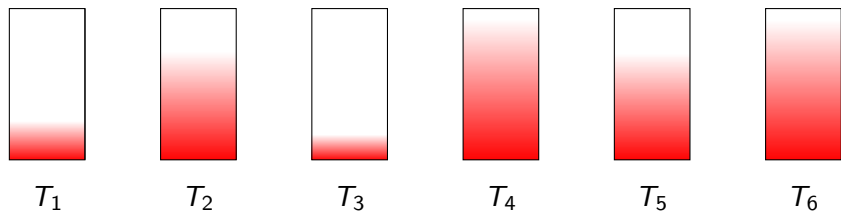


# Online VB: Step 6

Documents

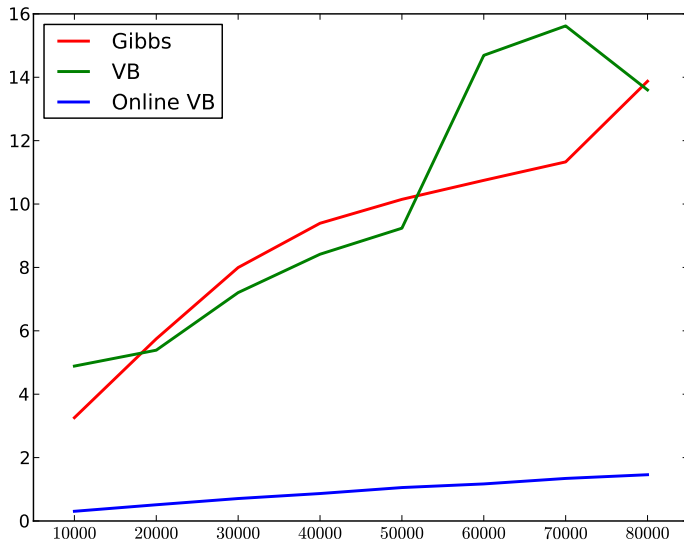


Topics

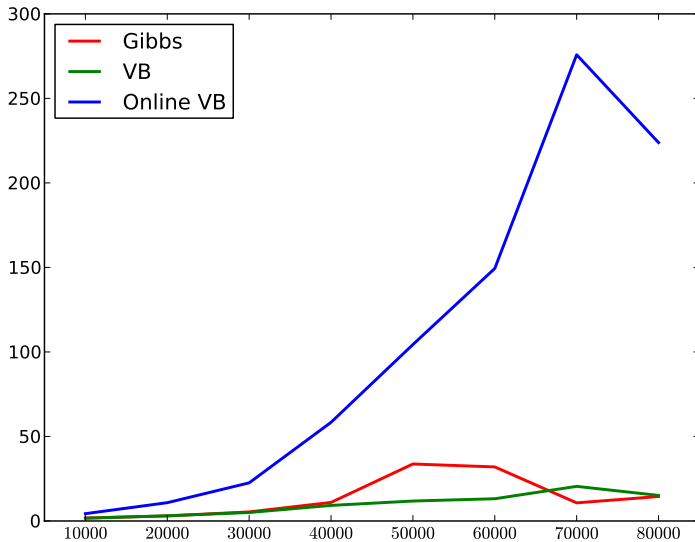


Comparison.

# Comparison: Time



# Comparison: Bound on Perplexity



Algorithms, live demo on  
<http://lda.ijs.si>