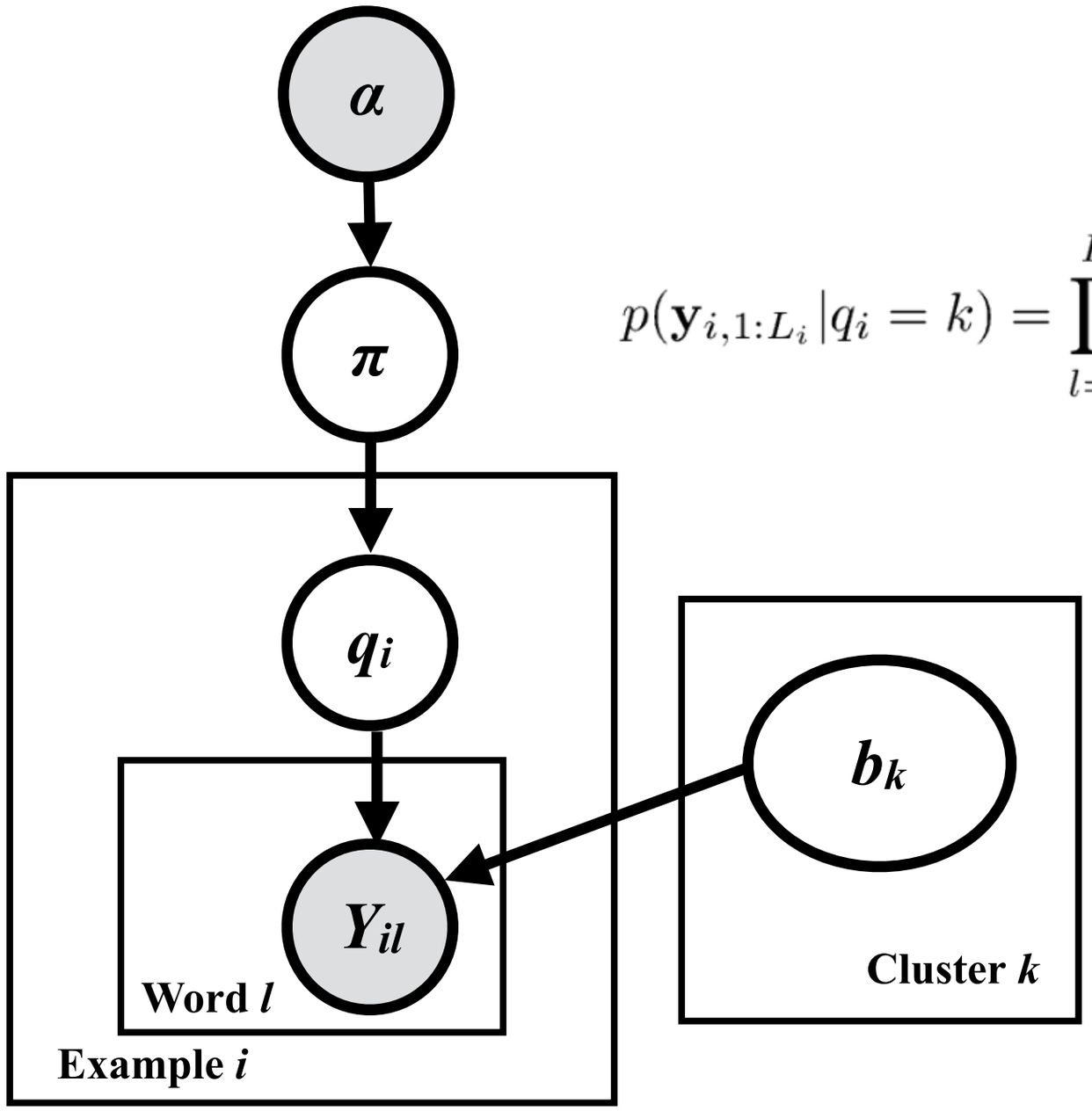
Chapter 27: Latent Dirichlet allocation

Text documents

Text document: List of words $\mathbf{y}_{i,1:L_i}$

Mixture model for discrete data



$$p(\mathbf{y}_{i,1:L_i}|q_i = k) = \prod_{l=1}^{L_i} \mathrm{Cat}(y_{il}|\mathbf{b}_k)$$

Different data structures require different parameterizations

List of words, same distribution per position:

$$p(\mathbf{y}_{i,1:L_i}|q_i = k) = \prod_{l=1}^{L_i} \mathrm{Cat}(y_{il}|\mathbf{b}_k)$$

Different distribution per position:

$$p(\mathbf{y}_{i,1:R}|q_i = k) = \prod_{r=1}^{R} \mathrm{Cat}(y_{il}|\mathbf{b}_k^{(r)})$$

Counts (fixed total):

$$p(\mathbf{n}_i|L_i, q_i = k) = \mathrm{Mu}(\mathbf{n}_i|L_i, \mathbf{b}_k)$$

Counts (unknown total):

$$p(\mathbf{n}_i|q_i = k) = \prod_{v=1}^{V} \mathrm{Poi}(n_{iv}|\lambda_{vk})$$

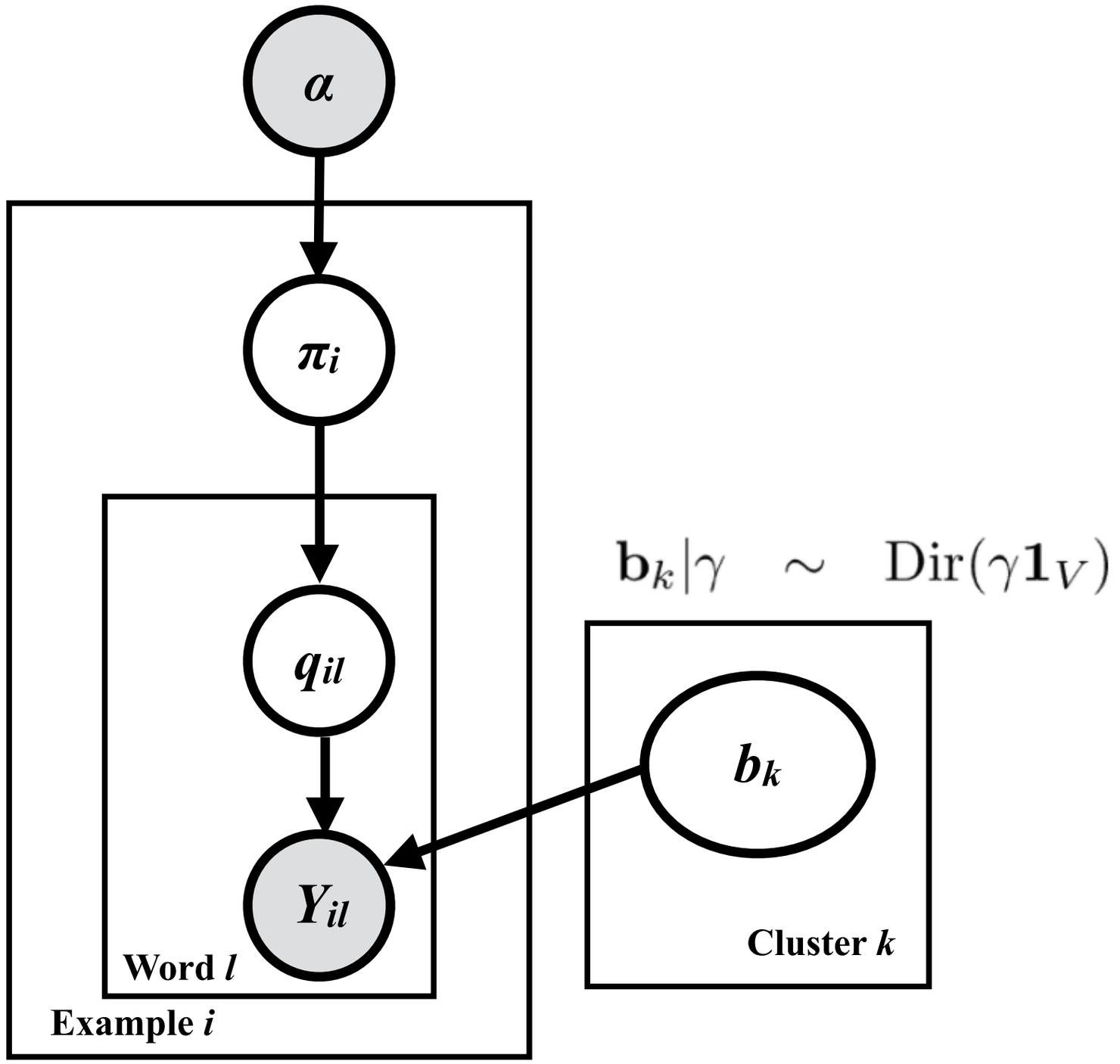
Latent Dirichlet allocation (LDA)

$$\boldsymbol{\pi}_i|\alpha \sim \mathrm{Dir}(\alpha \mathbf{1}_K)$$

$$q_{il}|\boldsymbol{\pi}_i \sim \mathrm{Cat}(\boldsymbol{\pi}_i)$$

$$y_{il}|q_{il}=k,\mathbf{B} \sim \mathrm{Cat}(\mathbf{b}_k)$$

$$\mathbf{b}_k|\gamma \sim \mathrm{Dir}(\gamma \mathbf{1}_V)$$



Latent Dirichlet allocation (LDA)

Topic 77

word	prob.
MUSIC	.090
DANCE	.034
SONG	.033
PLAY	.030
SING	.026
SINGING	.026
BAND	.026
PLAYED	.023
SANG	.022
SONGS	.021
DANCING	.020
PIANO	.017
PLAYING	.016
RHYTHM	.015
ALBERT	.013
MUSICAL	.013

Topic 82

word	prob.
LITERATURE	.031
POEM	.028
POETRY	.027
POET	.020
PLAYS	.019
POEMS	.019
PLAY	.015
LITERARY	.013
WRITERS	.013
DRAMA	.012
WROTE	.012
POETS	.011
WRITER	.011
SHAKESPEARE	.010
WRITTEN	.009
STAGE	.009

Topic 166

word	prob.
PLAY	.136
BALL	.129
GAME	.065
PLAYING	.042
HIT	.032
PLAYED	.031
BASEBALL	.027
GAMES	.025
BAT	.019
RUN	.019
THROW	.016
BALLS	.015
TENNIS	.011
HOME	.010
CATCH	.010
FIELD	.010

Document #29795

Bix beiderbecke, at age^{060} fifteen^{207}, sat^{174} on the slope^{071} of a bluff^{055} overlooking^{027} the mississippi^{137} river^{137}. He was listening^{077} to music^{077} coming^{009} from a passing^{043} riverboat. The music^{077} had already captured^{006} his heart^{157} as well as his ear^{119}. It was jazz^{077}. Bix beiderbecke had already had music^{077} lessons^{077}. He showed^{002} promise^{134} on the piano^{077}, and his parents^{035} hoped^{268} he might consider^{118} becoming a concert^{077} pianist^{077}. But bix was interested^{268} in another kind^{050} of music^{077}. He wanted^{268} to play^{077} the cornet. And he wanted^{268} to play^{077} jazz^{077}...

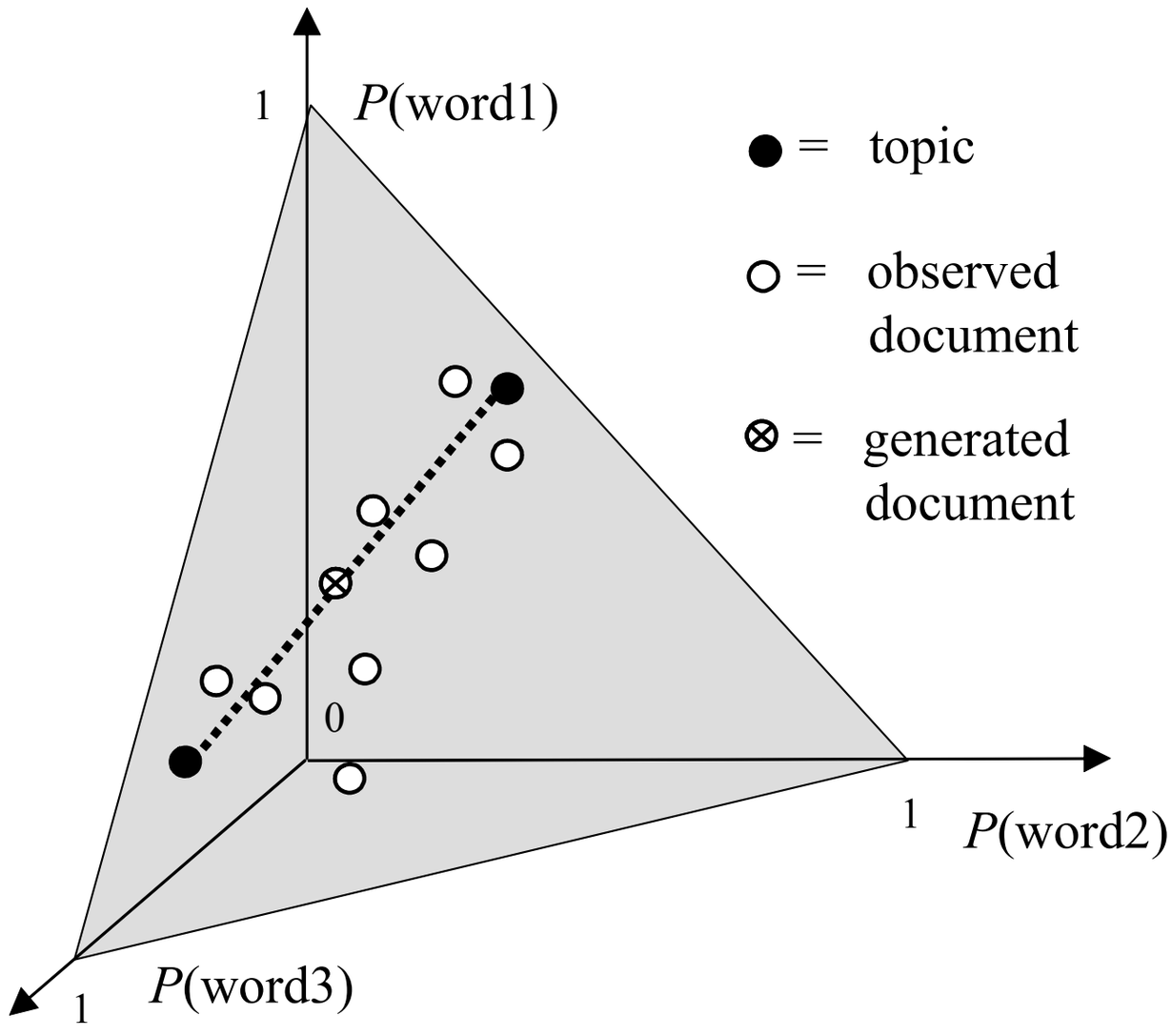
Document #1883

There is a simple^{050} reason^{106} why there are so few periods^{078} of really great theater^{082} in our whole western^{046} world. Too many things^{300} have to come right at the very same time. The dramatists must have the right actors^{082}, the actors^{082} must have the right playhouses, the playhouses must have the right audiences^{082}. We must remember^{288} that plays^{082} exist^{143} to be performed^{077}, not merely^{050} to be read^{254}. (even when you read^{254} a play^{082} to yourself, try^{288} to perform^{062} it, to put^{174} it on a stage^{078}, as you go along.) as soon^{028} as a play^{082} has to be performed^{082}, then some kind^{126} of theatrical^{082}...

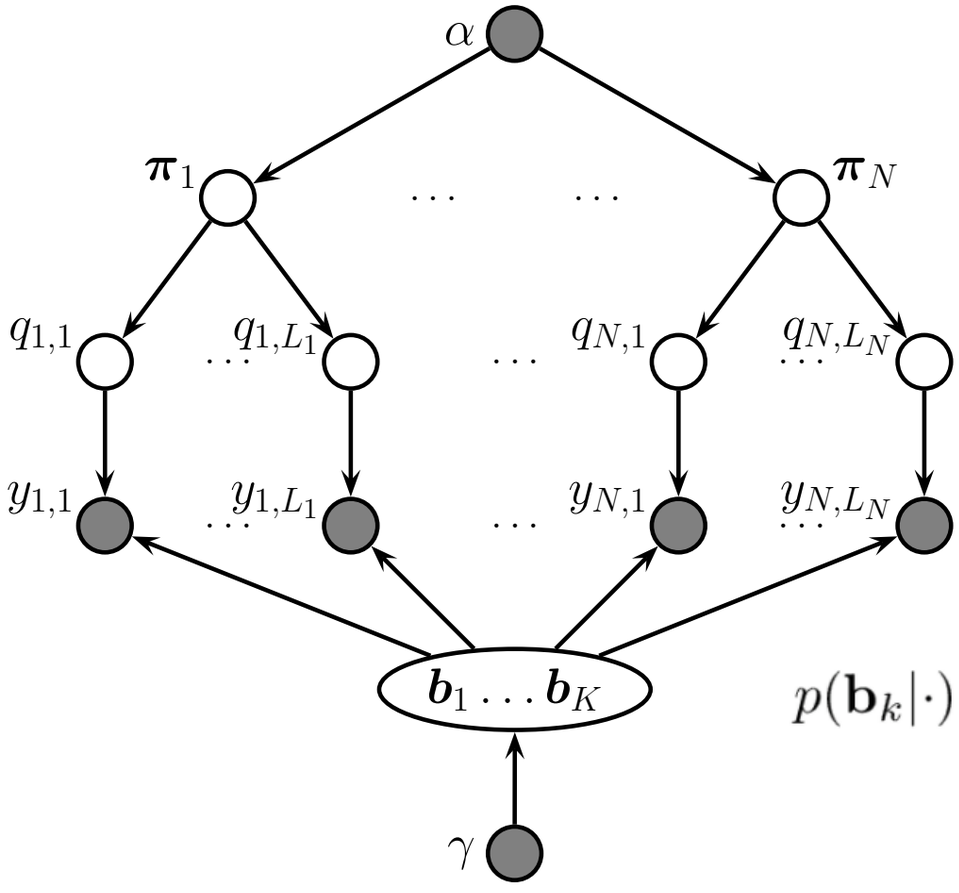
Document #21359

Jim^{296} has a game^{166} book^{254}. Jim^{296} reads^{254} the book^{254}. Jim^{296} sees^{081} a game^{166} for one. Jim^{296} plays^{166} the game^{166}. Jim^{296} likes^{081} the game^{166} for one. The game^{166} book^{254} helps^{081} jim^{296}. Don^{180} comes^{040} into the house^{038}. Don^{180} and jim^{296} read^{254} the game^{166} book^{254}. The boys^{020} see a game^{166} for two. The two boys^{020} play^{166} the game^{166}. The boys^{020} play^{166} the game^{166} for two. The boys^{020} like the game^{166}. Meg^{282} comes^{040} into the house^{282}. Meg^{282} and don^{180} and jim^{296} read^{254} the book^{254}. They see a game^{166} for three. Meg^{282} and don^{180} and jim^{296} play^{166} the game^{166}. They play^{166}...

LDA as dimensionality reduction



Fitting an LDA model using Gibbs sampling

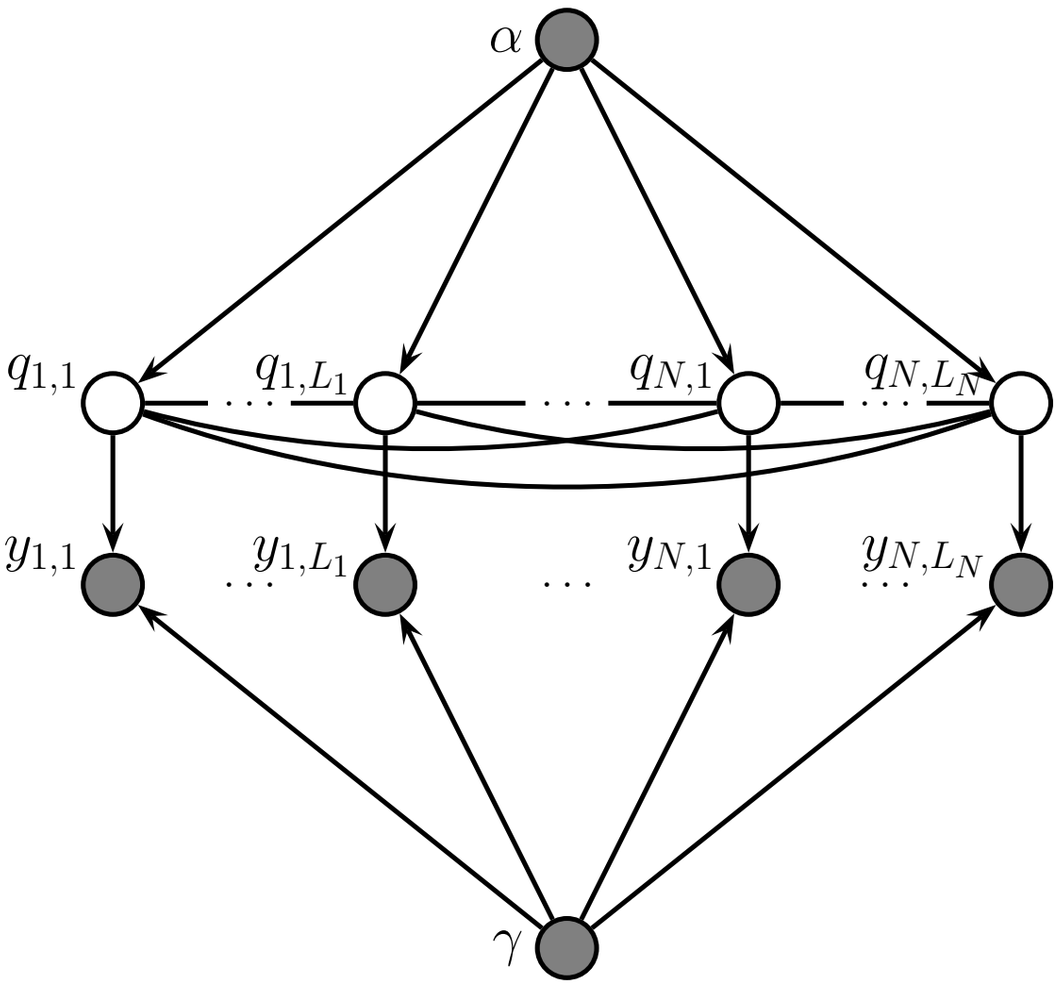


$$p(\boldsymbol{\pi}_i|\cdot) \quad = \quad \mathrm{Dir}(\{\alpha_k + \sum_l \mathbb{I}(z_{il} = k)\})$$

$$p(q_{il} = k|\cdot) \quad \propto \quad \exp[\log \pi_{ik} + \log b_{k,x_{il}}]$$

$$p(\mathbf{b}_k|\cdot) \quad = \quad \mathrm{Dir}(\{\gamma_v + \sum_i \sum_l \mathbb{I}(x_{il} = v, z_{il} = k)\})$$

Fitting an LDA model using collapsed Gibbs sampling



Exercise

We are interested in modeling the financial performance of companies across sectors. Each company's stock will either rise or fall in value in the next quarter. Each company has a particular sector *1...K* (agriculture, health, energy, etc). We observe *M* binary public attributes of each company (sales increasing/decreasing, public/private, etc).

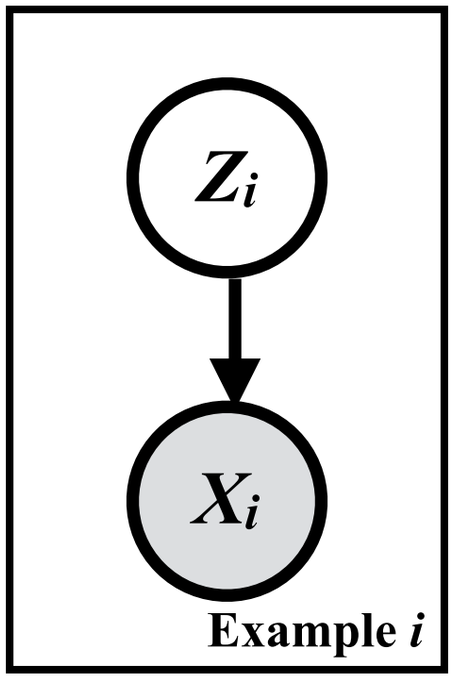
We will use a naive-Bayes-like model which imagines that the public attributes depend on both stock and sector, and that assumes that the attributes are independent of one another given stock and sector.

Draw a BN, MRF and factor graph that represent this model. Use plate notation. Propose a way to parameterize each distribution, and propose reasonable priors for each.

Exercise

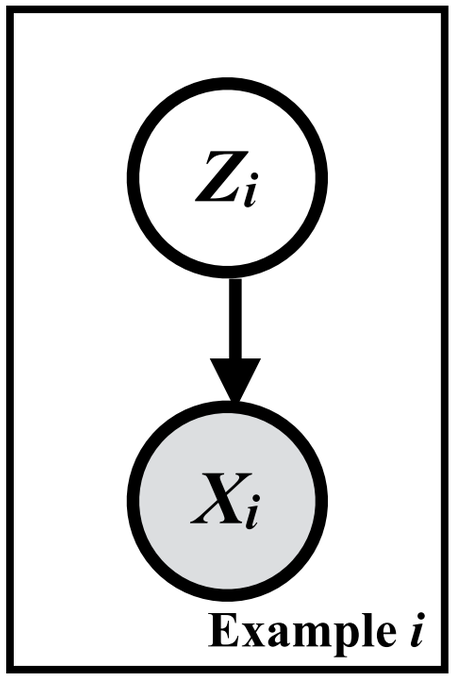
Assume stock and sector are observed for *N* companies. Derive the posterior distribution for all parameters.

Dimensionality reduction



$P(Z_i) \sim$ Normal

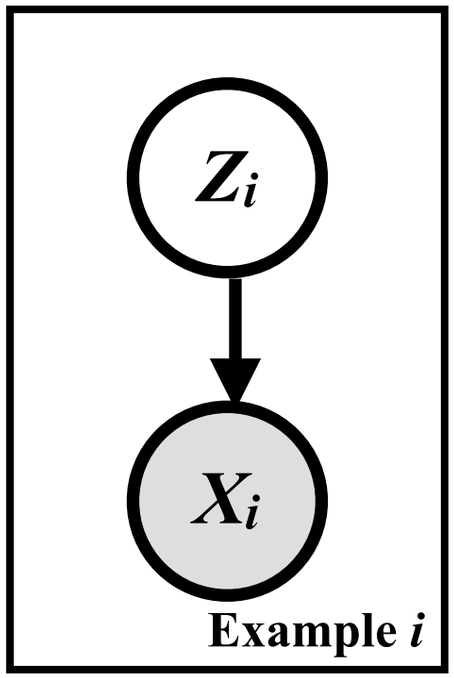
Variational approximation



$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \approx q_\lambda(z|x)$$

$$\text{maximize}_\lambda \quad KL(q_\lambda(z|x)||p(z|x))$$

Amortized inference

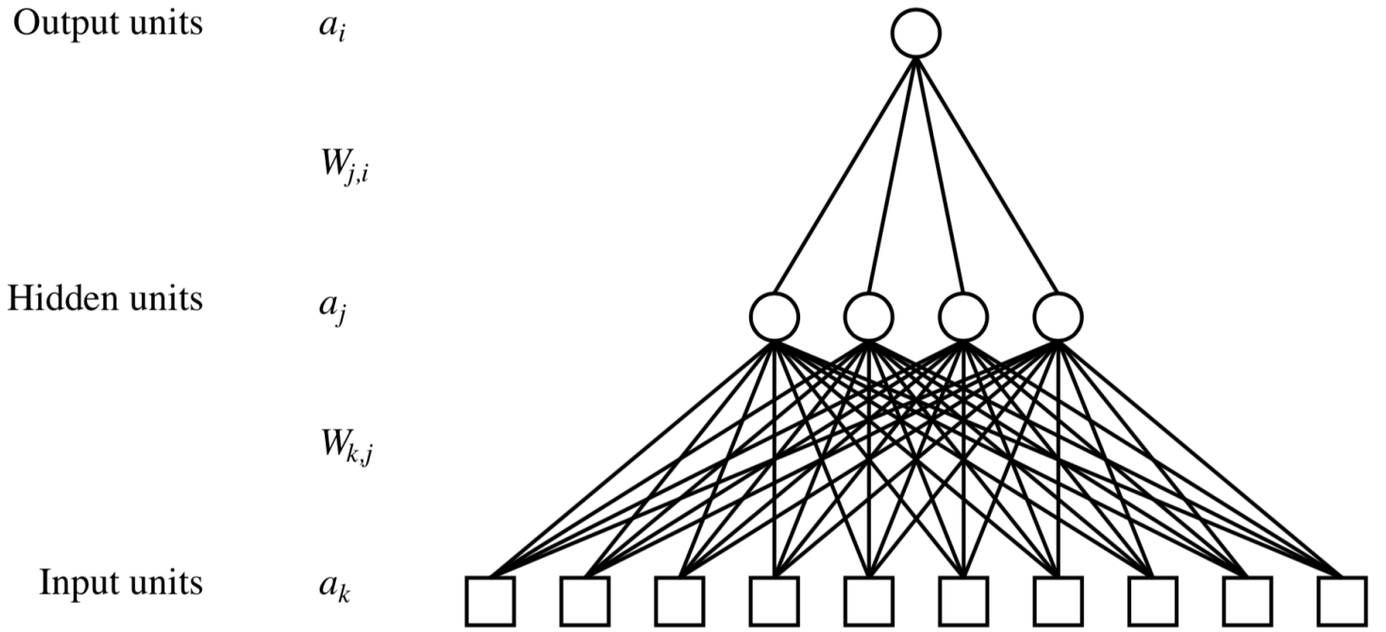


$$\text{maximize}_\lambda \quad KL(q_\lambda(z|x)||p(z|x))$$

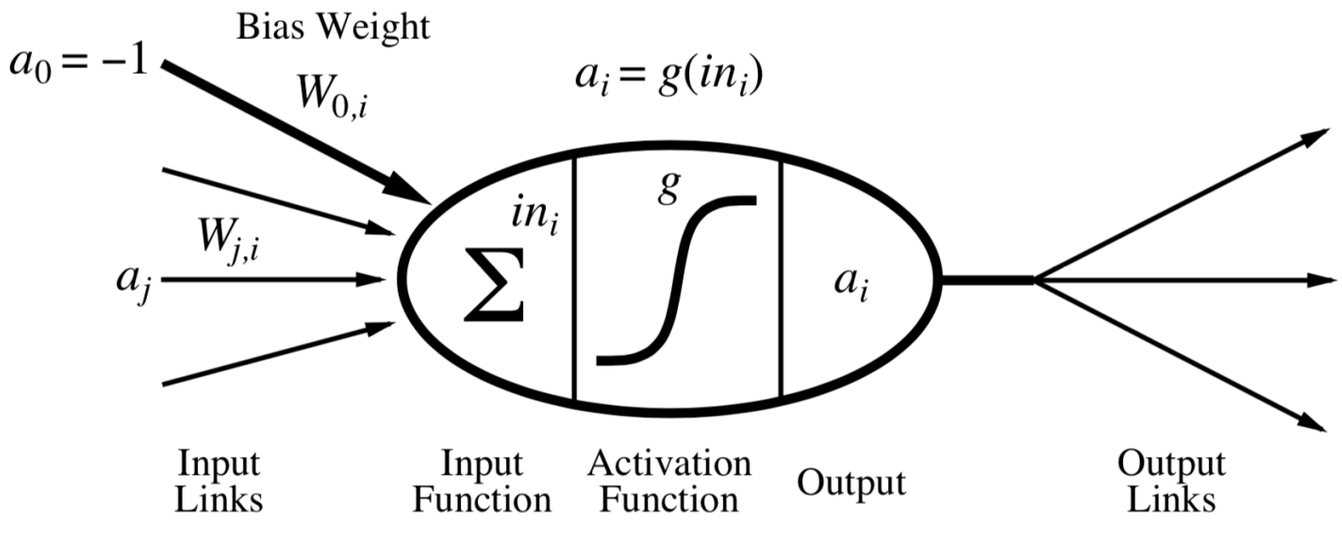
Idea #1: Use SGD

Idea #2: Make an algorithm $\lambda \leftarrow f(x)$ that works for most x.

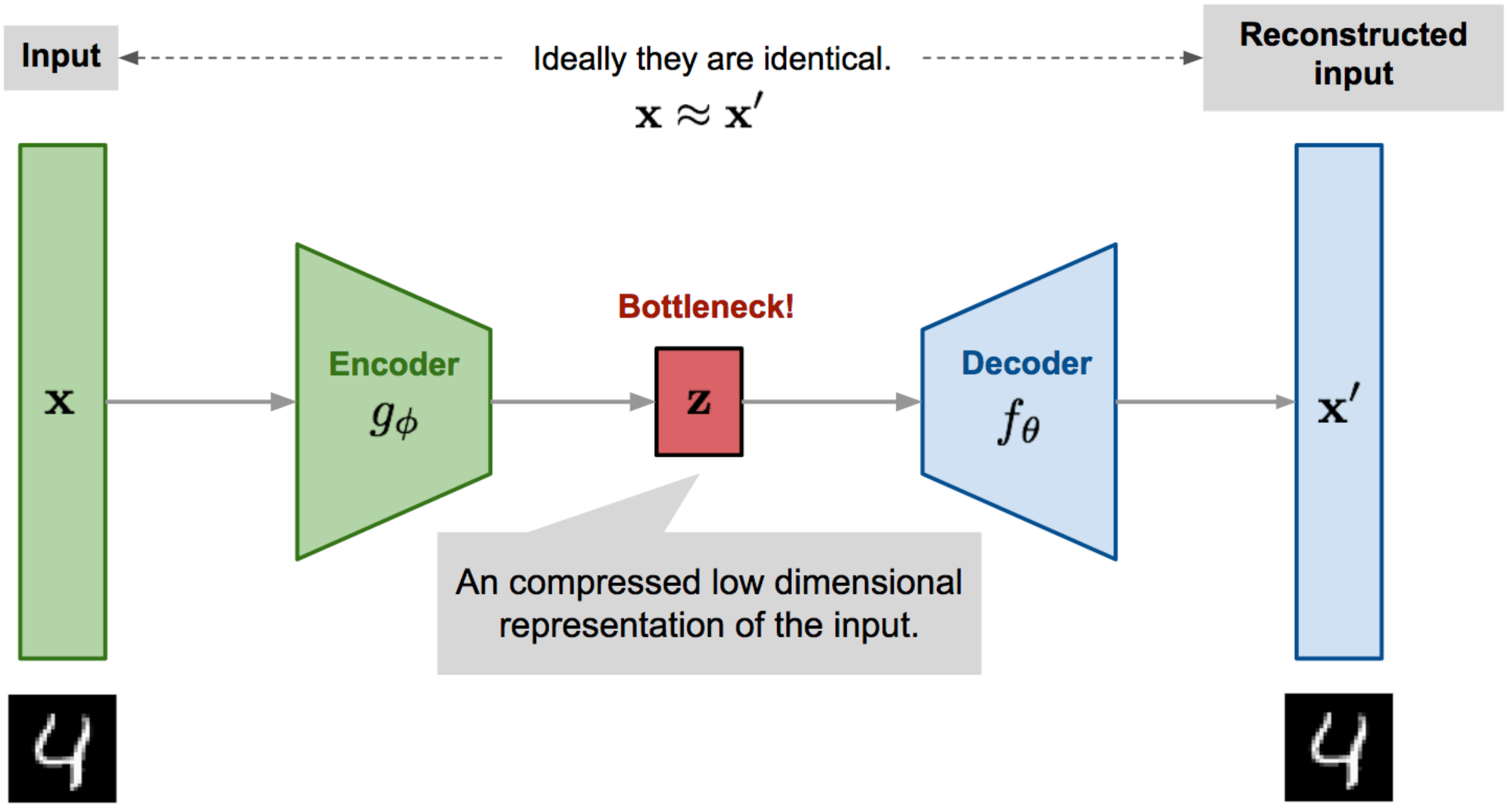
Neural network



$$a_i \leftarrow g(in_i) = g\left(\sum_j W_{j,i} a_j\right)$$



Autoencoder



Variational autoencoder