Review exercises

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).

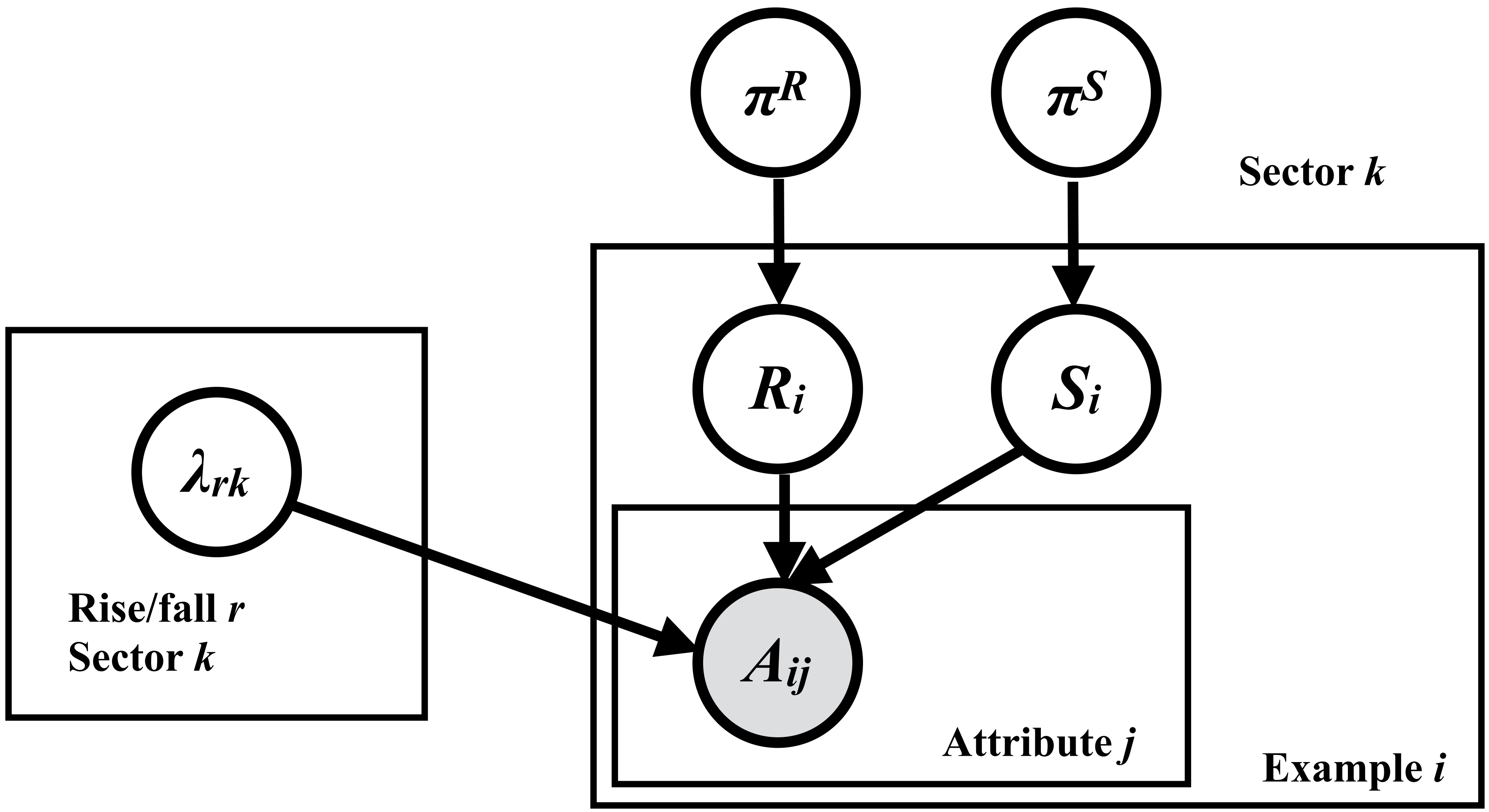
We will use a naive-Bayes-like model which imagines that the public attributes depend on both stock and sector, and that assumes that the attributes are independent of one another given stock and sector.

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.

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.



π^R
π^S
Sector *k*
R_i
S_i
λ_{rk}
Rise/fall *r*
Sector *k*
A_{ij}
Attribute *j*
Example *i*

Exercise

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

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

Exercise

Now we will assume that we never observe the stock or sector. We will build an unsupervised model. Of course, our learned clusters may or may not correspond to the known types.

Show how to use EM to learn the model parameters.

Now we will assume that we never observe the stock or sector. We will build an unsupervised model. Of course, our learned clusters may or may not correspond to the known types.

Show how to use EM to learn the model parameters.

Hint 1: Linearity of expectation

$$E_{p(x)}\left[\sum_i f(x_i)\right] = \sum_i E_{p(x)}[f(x_i)]$$

$$\begin{aligned}
E_{p(x)}\left[\sum_i f(x_i)\right] &= \sum_{\mathbf{x}} p(x) \sum_i f(x_i) \\
&= \sum_i \sum_{\mathbf{x}} p(x) f(x_i) \\
&= \sum_i \sum_{\mathbf{x_i}} \sum_{-\mathbf{x_i}} p(x_i) p(x_{-i}|x_i) f(x_i) \\
&= \sum_i \sum_{\mathbf{x_i}} p(x_i) f(x_i) \sum_{-\mathbf{x_i}} p(x_{-i}|x_i)
\end{aligned}$$

Hint 2

$$\mathrm{argmax}_\pi \sum_k \alpha_k \log \pi_k = \mathrm{argmax}_\pi E_\alpha[\log \pi_k] = \frac{1}{Z}\alpha$$

Hint 2

$$\operatorname{argmax}_\pi \sum_k \alpha_k \log \pi_k = \operatorname{argmax}_\pi E_\alpha[\log \pi_k] = \frac{1}{Z}\alpha$$

$$\begin{aligned}
&\operatorname{argmax}_\pi \sum_k \alpha_k \log \pi_k \\
&= \operatorname{argmax}_\pi \sum_k \alpha_k \log \pi_k - \sum_k \alpha_k \log \alpha_k \\
&= \operatorname{argmax}_\pi \sum_k \alpha_k \log \frac{\pi_k}{\alpha_k} \\
&= \operatorname{argmax}_\pi - KL(\alpha \| \pi)
\end{aligned}$$

Hint 3: EM for a mixture model

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}) &\triangleq \mathbb{E}\left[\sum_i \log p(\mathbf{x}_i, z_i|\boldsymbol{\theta})\right] \\
&= \sum_i \mathbb{E}\left[\log\left[\prod_{k=1}^{K} (\pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k))^{\mathbb{I}(z_i=k)}\right]\right] \\
&= \sum_i \sum_k \mathbb{E}\left[\mathbb{I}(z_i = k)\right] \log[\pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k)] \\
&= \sum_i \sum_k p(z_i = k|\mathbf{x}_i, \boldsymbol{\theta}^{t-1}) \log[\pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k)] \\
&= \sum_i \sum_k r_{ik} \log \pi_k + \sum_i \sum_k r_{ik} \log p(\mathbf{x}_i|\boldsymbol{\theta}_k)
\end{aligned}
$$

$$
r_{ik} = \frac{\pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k^{(t-1)})}{\sum_{k'} \pi_{k'} p(\mathbf{x}_i|\boldsymbol{\theta}_{k'}^{(t-1)})}
$$

Solution

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A, S, P|\theta^{(t-1)}) \log P(A, S, P|\theta)$$

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A, S, P|\theta^{(t-1)}) \log P(A, S, P|\theta)$$

$$\log P(A, S, P|\theta) = \sum_{i} \log P(S_i)P(P_i)P(A_i|S_i, P_i)$$

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A, S, P|\theta^{(t-1)}) \log P(A, S, P|\theta)$$

$$\log P(A, S, P|\theta) = \sum_{i} \log P(S_i)P(P_i)P(A_i|S_i, P_i)$$

$$= \sum_{i} \log \pi^{S}_{S_i} + \log \pi^{P}_{P_i} + \sum_{j} \log \lambda_{j,S_i,P_i}$$

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A, S, P|\theta^{(t-1)}) \log P(A, S, P|\theta)$$

$$\log P(A, S, P|\theta) = \sum_i \log P(S_i)P(P_i)P(A_i|S_i, P_i)$$

$$= \sum_i \log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i}$$

$$Q(\theta, \theta^{(t-1)}) = \sum_i p(S_i, P_i, A_i|\theta^{(t-1)}) \left[\log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i}\right]$$

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A, S, P|\theta^{(t-1)}) \log P(A, S, P|\theta)$$

$$\log P(A, S, P|\theta) = \sum_i \log P(S_i) P(P_i) P(A_i|S_i, P_i)$$

$$= \sum_i \log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i}$$

$$Q(\theta, \theta^{(t-1)}) = \sum_i p(S_i, P_i, A_i|\theta^{(t-1)}) \left[\log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i} \right]$$

$$r_{i,j,k} = P(S_i = j, P_i = k, A_i|\theta^{(t-1)}) = \pi_j^{S,(t-1)} \pi_k^{P,(t-1)} \prod_j \lambda^{(t-1)}_{j,S_i,P_i}$$

Solution

$$Q(\theta, \theta^{(t-1)}) = \sum_{S,P} p(A,S,P|\theta^{(t-1)}) \log P(A,S,P|\theta)$$

$$\log P(A,S,P|\theta) = \sum_i \log P(S_i)P(P_i)P(A_i|S_i,P_i)$$

$$= \sum_i \log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i}$$

$$Q(\theta, \theta^{(t-1)}) = \sum_i p(S_i, P_i, A_i|\theta^{(t-1)}) \left[\log \pi^S_{S_i} + \log \pi^P_{P_i} + \sum_j \log \lambda_{j,S_i,P_i}\right]$$

$$r_{i,j,k} = P(S_i = j, P_i = k, A_i|\theta^{(t-1)}) = \pi_j^{S,(t-1)} \pi_k^{P,(t-1)} \prod_j \lambda^{(t-1)}_{j,S_i,P_i}$$

$$Q(\lambda_{j,k,l}, \theta^{(t-1)}) \propto KL\left(\sum_i r_{i,k,l} 1(A_i = j) \| \lambda_{j,k,l}\right)$$

$$Q(\pi_k, \theta^{(t-1)}) \propto KL\left(\sum_{i,l} r_{i,j,k} \| \lambda_{j,k,l}\right)$$

Exercise

1) Describe how to use Gibbs sampling for inference in this model. Derive the update probabilities.

2) Suppose we use the Metropolis-Hastings for inference, using a transition that chooses both R_i and S_i (simultaneously) uniformly at random. Derive the acceptance probability of this transition.

$$\alpha = \frac{p^*(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{p^*(\mathbf{x})q(\mathbf{x}'|\mathbf{x})} = \frac{p^*(\mathbf{x}')/q(\mathbf{x}'|\mathbf{x})}{p^*(\mathbf{x})/q(\mathbf{x}|\mathbf{x}')}$$

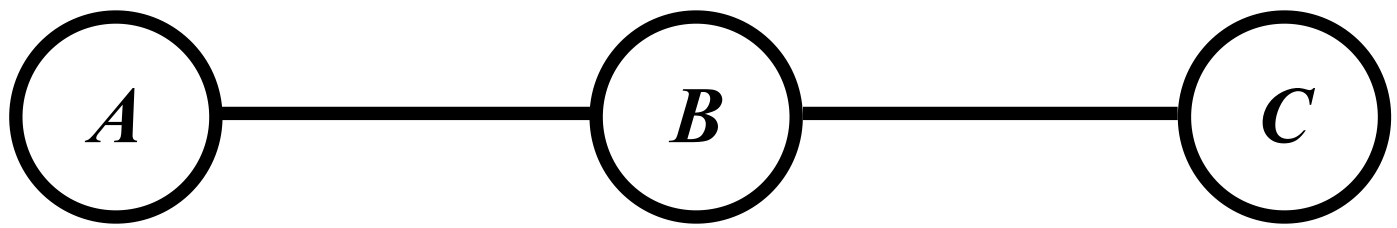
1) Describe how to use Gibbs sampling for inference in this model. Derive the update probabilities.

2) Suppose we use the Metropolis-Hastings for inference, using a transition that chooses both R_i and S_i (simultaneously) uniformly at random. Derive the acceptance probability of this transition.

$$\alpha = \frac{p^*(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{p^*(\mathbf{x})q(\mathbf{x}'|\mathbf{x})} = \frac{p^*(\mathbf{x}')/q(\mathbf{x}'|\mathbf{x})}{p^*(\mathbf{x})/q(\mathbf{x}|\mathbf{x}')}$$

Exercise

Consider the following model with binary variables *A, B* and *C*. Compute a mean field update for B in the following model for iteration *t+1*.



$$\psi_{A,B} = \psi_{B,C} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$q_A^t(0) = 1/4$$
$$q_B^t(0) = 2/3$$

Consider the following model with binary variables *A, B* and *C*. Compute a mean field update for B in the following model for iteration *t+1*.

$$\log q_j(x_j) = E_{x_{-j} \sim q_{-j}}[\log p(x)] = \sum_{x_{-j}} \prod_{k \neq j} q_k(x_k) \log p(x)$$

$$A - B - C$$

$$\psi_{A,B} = \psi_{B,C} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$q_A^t(0) = 1/4$$

$$q_B^t(0) = 2/3$$

Exercise

Consider the probability distribution

$$p(x) = \frac{1}{Z}\exp(-x)$$

1) Determine the value of the normalization constant Z.

2) Show how to use the inverse CDF method to sample *x* from this distribution.

Consider the probability distribution $$p(x) = \frac{1}{Z}\exp(-x)$$

1) Determine the value of the normalization constant Z.

2) Show how to use the inverse CDF method to sample *x* from this distribution.

Statistical machine learning