CDS: Machine Learning WK4-7

Jaqueline Zeitler Sander Meis Wieke Bergers

August 2020

Week 6

Graphical models

a

$$\begin{aligned} p(x_i|y) =& \frac{p(x_i, y)}{p(y)} \\ =& \frac{1}{p(y)} \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} p(x_1, x_2, \ldots, x_n, y) \\ =& \frac{1}{p(y)} \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} p(x_1)p(x_2)\ldots p(x_n)p(y|x_1, x_2, \ldots, x_n) \end{aligned}$$

b

$$p(x_i|y, x_1 = 1, x_1 = 1, i \neq 1) = \frac{1}{p(y)} \sum_{x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_n} p(x_1)p(x_2)\ldots p(x_n)p(y|x_1, x_2, \ldots, x_n)$$

a

Mean is simple:

$$\mu = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \mu_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \gamma(x_1 + x_2) \end{pmatrix}$$

Because $x1$ and x_2 are independent, the upper left covariance matrix is just a diagonal matrix:

$$\mathbf{\Sigma}_{x_1,x_2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{aligned} \Sigma_{13} &= \text{cov}(x_1, y) = E(x_1 y) - E(x_1)E(y) = E(x_1)(\gamma(x_1 + x_2) + \chi)) \\ &= E(x_1(\gamma(x_1 + x_2))) + E(x_1\chi) = E(\gamma x_1^2) + E(\gamma x_1 x_2) = \gamma \end{aligned}$$

As y depends in the same way on x_1 and x_2, the covariances of y with either of them is the same. The variance of y is given by

$$E(y^2) - E(y)^2 = E((\gamma x_1 + \gamma x_2 + \chi)^2) - E(\gamma(x_1 + x_2))^2$$
$$= E(\gamma^2 x_1^2 + 2\gamma^2 x_1 x_2 + 2\gamma x_1 \chi + \gamma^2 x_2^2 + 2\gamma x_2 \chi + \chi^2) - 0^2$$
$$= \gamma^2 E(x_1^2 + 2x_1 x_2 + x_2^2) + E(2\gamma x_1 \chi + 2\gamma x_2 \chi + \chi^2)$$
$$= 2\gamma^2 + E(\chi^2) = 2\gamma^2 + \sigma^2$$

So we get for the covariance matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & \gamma \\ 0 & 1 & \gamma \\ \gamma & \gamma & 2\gamma^2 + \sigma^2 \end{pmatrix}$$

Since the probability is a Gaussian:

$$p(x_1, x_2, y) = \mathcal{N}(x_1, x_2, y | \mu, \mathbf{\Sigma})$$

b

$$\mu_{\mathbf{x_1}, \mathbf{x_2} | \mathbf{y}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} \frac{\chi}{2\gamma^2 + \sigma^2} = \frac{\gamma\chi}{2\gamma^2 + \sigma^2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\mathbf{\Sigma}_{\mathbf{x_1}, \mathbf{x_2} | \mathbf{y}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} \frac{1}{2\gamma^2 + \sigma^2} \begin{pmatrix} \gamma & \gamma \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{\gamma^2}{2\gamma^2 + \sigma^2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} & -\frac{\gamma^2}{2\gamma^2+\sigma^2} \\ -\frac{\gamma^2}{2\gamma^2+\sigma^2} & 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} \end{pmatrix}$$

c

$$\rho = \frac{-\frac{\gamma^2}{2\gamma^2+\sigma^2}}{1 - \frac{\gamma^2}{2\gamma^2+\sigma^2}} = \frac{-\gamma^2}{2\gamma^2 + \sigma^2 - \gamma^2} = \frac{-\gamma^2}{\sigma^2 + \gamma^2}$$

Week 7

EM 2.3

1.

See python notebook

2.

Gaussian mixture model:

$$p(x,k) = \pi_k \mathcal{N}(x \mid a_k, \Sigma_k) \tag{2}$$

the logarithm of a Gaussian is

$$\log \mathcal{N}(x \mid a_k, \Sigma_k) = -\frac{(x-a_k)^2}{2\sigma_k^2} - \log \sigma_k - \frac{\log 2\pi}{2} \tag{4}$$

the responsibility is then

$$r_{\mu k} = \frac{\pi_k \mathcal{N}(x^\mu \mid \mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} \mathcal{N}(x^\mu \mid \mu_{k'}, \Sigma_{k'})} \tag{6}$$

the variational bound is then

$$Q(\theta, q^*) = L(\theta) = \sum_\mu \sum_k D(x) r_{\mu k} \log \frac{p(x,k \mid \theta)}{r_{\mu k}} \tag{8}$$

$$Q(\theta) = \sum_\mu D(x) \sum_k r_{\mu k} \log p(x,k \mid \theta) \tag{9}$$

$$Q(\theta) = \sum_\mu \sum_k r_{\mu k} \left(\log \pi_k + \log \mathcal{N}(x^\mu \mid \theta)\right) \tag{10}$$

next we approximate that the responsibility is a constant with regards to a_k and σ_k, use Lagrange multipliers and derive for a specific k

$$\sum_k P_k = 1 \tag{12}$$

$$Q(\theta_k) = \sum_\mu \sum_k r_{\mu k} \left(\log \pi_k + \log \mathcal{N}(x^\mu \mid \theta)\right) - \lambda(\sum_k \pi_k - 1) \tag{13}$$

$$\frac{\partial Q}{\partial \pi_k} = \sum_\mu \frac{r_{\mu k}}{\pi_k} - \lambda = 0 \tag{14}$$

$$\pi_k = \sum_\mu r_{\mu k} / \lambda \tag{15}$$

now we find λ by optimizing

$$ \tag{16} $$

$$\frac{\partial Q}{\partial \lambda} = -\sum_k \pi_k + 1 = 0 \tag{17}$$

$$\sum_k \pi_k = 1 = \sum_k \sum_\mu r_{\mu k}/\lambda \tag{18}$$

$$\lambda = \sum_k \sum_\mu r_{\mu k} \tag{19}$$

$$\pi_k = \frac{\sum_\mu r_{\mu k}}{\sum_k \sum_\mu r_{\mu k}} \tag{20}$$

$$r_k \equiv \sum_\mu r_{\mu k} \tag{21}$$

$$N \equiv \sum_k r_k \tag{22}$$

$$\pi_k = \frac{r_k}{N} \tag{23}$$

now for a_k

$$\frac{\partial Q}{\partial a_k} = \sum_\mu r_{\mu k} \frac{x^\mu - a_k}{\sigma_k^2} = 0 \tag{24}$$

$$a_k \sum_\mu r_{\mu k} = \sum_\mu r_{\mu k} x^\mu \tag{25}$$

$$a_k = \frac{\sum_\mu r_{\mu k} x^\mu}{\sum_\mu r_{\mu k}} \tag{26}$$

$$a_k = \frac{\sum_\mu r_{\mu k} x^\mu}{r_k} \tag{27}$$

now for σ_k

$$\tag{28}$$

$$\frac{\partial Q}{\partial \sigma_k} = \sum_{\mu} r_{\mu k} \left(\frac{(x^\mu - a_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right) = 0 \tag{29}$$

$$\frac{\partial L}{\partial \sigma_k} = \sum_{\mu} r_{\mu k}(x^\mu - a_k)^2 - r_{\mu k}\sigma_k^2 = 0 \tag{30}$$

$$\sigma_k^2 \sum_{\mu} r_{\mu k} = \sum_{\mu} r_{\mu k}(x^\mu - a_k)^2 \tag{31}$$

$$\sigma_k^2 = \frac{\sum_\mu r_{\mu k}(x^\mu - a_k)^2}{\sum_\mu r_{\mu k}} \tag{32}$$

$$\sigma_k^2 = \frac{\sum_\mu r_{\mu k}(x^\mu - a_k)^2}{r_k} \tag{33}$$

now we fill in for a_k

$$\tag{34}$$

$${\sigma_k}^2 = \frac{\sum_\mu r_{\mu k}\left({x^\mu}^2 - 2x^\mu a_k + {a_k}^2\right)}{r_k} \tag{35}$$

$${\sigma_k}^2 = \frac{\sum_\mu r_{\mu k}{x^\mu}^2}{r_k} - \frac{\sum_\mu 2r_{\mu k}x^\mu a_k}{r_k} + \frac{\sum_\mu r_{\mu k}{a_k}^2}{r_k} \tag{36}$$

$${\sigma_k}^2 = \frac{\sum_\mu r_{\mu k}{x^\mu}^2}{r_k} - 2{a_k}^2 + {a_k}^2 \tag{37}$$

$${\sigma_k}^2 = \frac{\sum_\mu r_{\mu k}{x^\mu}^2}{r_k} - {a_k}^2 \tag{38}$$

This makes sense compared to the multidimensional case from the slides, when $i == j$ we get the previous result.