COMP9418 Assignment 1

Advanced Topics in Statistical Machine Learning, 17s2, UNSW Sydney

Last Update: Wednesday 23$^{\text{rd}}$ August, 2017 at 10:59

Submission deadline: Wednesday September 6th, 2017 at 23:59:59

Late Submission Policy: 20% marks will be deducted from the total for each day late, up to a total of four days. If five or more days late, a zero mark will be given.

Form of Submission: You should submit your solution with the following files:

1. `solution.pdf`: Theory part;
2. `solution.ipynb`: Jupyter notebook; and
3. `model.npz`: The model in compressed .npz format.

No other formats will be accepted (scanned versions of legible handwritten answers are accepted for the theory part). There is a maximum file size cap of 20MB so make sure your submission does not exceed this size.

Submit your files using give. On a CSE Linux machine, type the following on the command-line:

```
$ give cs9418 ass1 solution.pdf solution.ipynb model.npz
```

Alternative, you can submit your solution via the course website
`https://webcms3.cse.unsw.edu.au/COMP9418/17s2/resources/12151`

Recall the guidance regarding plagiarism in the course introduction: this applies to this homework and if evidence of plagiarism is detected it may result in penalties ranging from loss of marks to suspension.

1 [50 Marks] Mixture Models

Consider the following model for categorical variable $z \in \{1, \ldots, K_z\}$ and D-dimensional vector $\mathbf{x}$ where each $x_d \in \{1, \ldots, K_x\}$:

$$p(z|\boldsymbol{\pi}) = \mathrm{Cat}(z|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{\mathbb{I}(z=k)}, \text{ where } \boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^T,\ \pi_k \geq 0,\ \sum_{k=1}^{K} \pi_k = 1, \tag{1}$$

$$p(\mathbf{x}|z, \boldsymbol{\phi}) = p(x_1|z, \boldsymbol{\phi}_1) p(x_2|x_1, z, \boldsymbol{\phi}_2) \prod_{d=3}^{D} p(x_d|x_{d-2}, x_{d-1}, z, \boldsymbol{\phi}_d), \tag{2}$$

where the conditional distributions for $x_1, \ldots, x_d$ are also Categorical distributions, with parameters $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_D\}$, and we note that $\{\boldsymbol{\phi}_d\}_{d=1}^{D}$ are vectors themselves. Denoting the observations $\mathbf{X} = \{\mathbf{x}^{(n)}\}_{n=1}^{N}$, corresponding latent variables $\mathbf{z} = \{z^{(n)}\}_{n=1}^{N}$, and making i.i.d. assumptions on the prior over latent variables and the conditional likelihood we have that:

$$p(\mathbf{z}|\boldsymbol{\pi}) = \prod_{n=1}^{N} p(z^{(n)}|\boldsymbol{\pi}), \text{ and } p(\mathbf{X}|\mathbf{z}, \boldsymbol{\phi}) = \prod_{n=1}^{N} p(\mathbf{x}^{(n)}|z^{(n)}, \boldsymbol{\phi}). \tag{3}$$

Denote the joint distribution with $p(\mathbf{X}, \mathbf{z}|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\phi}\}$ are the complete set of model parameters.

1. [10 Marks] Using $D = 5$, draw the graphical model for the joint distribution $p(\mathbf{X}, \mathbf{z}|\boldsymbol{\theta})$ above (do not include parameters $\boldsymbol{\theta}$ in the graph).

2. [10 Marks] For general D, K_z, K_x, N, how many parameters do we need to estimate in this model?

3. [10 Marks] Write down the full expression for the expected complete-data log likelihood (also known as auxiliary function) for this model, i.e. $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \mathbb{E}_{p(\mathbf{z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})}[\log p(\mathbf{X}, \mathbf{z}|\boldsymbol{\theta})]$. Show all your working.

4. [10 Marks] Use the expression of the expected complete-data log likelihood to derive Expectation Maximization (EM) updates for the parameters of the model $\boldsymbol{\theta}$. Show all your working and highlight each individual update equation.

5. [10 Marks] Give the computational complexity per iteration of this specific EM algorithm in terms of N, D, K_z, K_x.

2 [50 Marks] Practical Part

See Jupyter notebook `comp9418_ass1.ipynb`.