Forecast and Simulation **Summer Term 2023**
http://ml3.leuphana.de/lectures/summer23/FS/
Machine Learning Group, Leuphana University of Lüneburg

Ulf Brefeld (brefeld@leuphana.de)
Yannick Rudolph (yannick.rudolph@leuphana.de)

Exercise 7

Due: Thursday, June 15, 2023

Task 1: Mixtures of Bernoulli distributions

This copy of Task 3 from Exercise 6 builds upon the exposition in Section 9.3.3 of [Bishop, 2006].

i) Show that if we want to maximize the expected complete-data log likelihood function (Equation 9.55 in [Bishop, 2006]) for a mixture of Bernoulli distributions wrt. the mixing coefficients π_k (using a Lagrange multiplier to enforce the summation constraint), we obtain Equation 9.60.

ii) Reimplement the experiment described in Section 9.3.3 of [Bishop, 2006] (and illustrated in Figure 9.10). You will have to implement the EM algorithm for mixtures of Bernoulli distributions. We provide a fixed dataset in *exercise-06-EM.ipynb* (which involves downloading MNIST).

Answer

i) See classnotes.

ii) See *solution-07-EM.ipynb*.

Task 2: Expectation Maximization (continued)

Please also read about *The EM Algorithm in General* in Section 9.4 of [Bishop, 2006].

i) Proof that

$$\ln p(X|\theta) = \mathcal{L}(q, \theta) + \mathrm{KL}(q\|p),$$

i.e. Equation 9.70 of [Bishop, 2006], where $\mathcal{L}(q, \theta)$ and $\mathrm{KL}(q\|p)$ are defined in Equations 9.71 and 9.72, respectively.

ii) Why can we consider $\mathcal{L}(q, \theta)$ a lower bound on $\ln p(X|\theta)$?

iii) In an alternative derivation, show that $\ln p(X|\theta) \geq \mathcal{L}(q, \theta)$ by means of Jenssen's inequality.

iv) Please show that the gradient of $\mathcal{L}(q, \theta)$ wrt. θ is equal to the gradient of $\ln p(X|\theta)$ at $\theta = \theta^{\text{old}}$ if we set $q(Z) = p(Z|X, \theta^{\text{old}})$.

v) For Equation 9.75 of [Bishop, 2006], please show that the last equality holds.

Answer

See classnotes.

References

[Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.