Expectation Maximization

HU, Pili*

April 15, 2012†

Abstract

You're not reading the completed document. Only brainstorm and some message from the author are there.

*hupili [at] ie [dot] cuhk [dot] edu [dot] hk

†Last compile:April 15, 2012

Contents

1 Brainstorm **3**

Acknowledgements **3**

References **3**

Appendix **3**

1 Brainstorm

Topics to becovered:

- Collect how different people reach the Q function, like [2][1][6]. Some are later interpretations, but worth to know.
- Why do we choose ln in the Q function? Are there any other possibilities? I just find that any concave function with $f(1) = 0$ suffices in the derivation of general form. The benefit of ln simply lies in the fact that it favours the exponential family.
- An application of GMM.

Acknowledgements

References

[1] C.M Bishop. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[2] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, 1977.

[3] Pili Hu. Expectation maximization tutorial. GitHub, https://github.com/hupili/tutorial/tree/master/expectation-maximization, 4 2012. HU, Pili's tutorial collection.

[4] Pili Hu. Tutorial collection. GitHub, https://github.com/hupili/tutorial, 3 2012. HU, Pili's tutorial collection.

[5] Expectation Maximization Algorithm, Wikipedia, `http://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm`

[6] Sean Borman, The Expectation Maximization Algorithm: A short tutorial, 2004, `http://www.seanborman.com/publications/EM_algorithm.pdf`

Appendix

Message From the Author

The first time heard of EM algorithm was about 3 years ago. During all these years, I tried to learn and re-learn it for many times. Every time

I find something new. Theoretists and practitioners have different tastes in explaining things. The former makes rigorous derivations and provides different views. The latter constructs guide rules for people who only wants a solution but doesn't bother to derive.

Recently, I eventually managed to derive the EM algorithm in general and apply it to the GMM Maximum Likelihood learning problem. Although there are many tutorials out there, I determine to write my own, making it complete with theory, algorithm, and implementation. More importantly, I make the source publicly open. If people want to fix anything or make their own notes upon this document, they should find it convenient.