Naturally Constrained Online Expectation Maximization

Daniela Pamplona, Antoine Manzanera

U2IS, ENSTA Paris, Institut Polytechnique de Paris,
828 Boulevard des Maréchaux,
91120 Palaiseau, France

Introduction

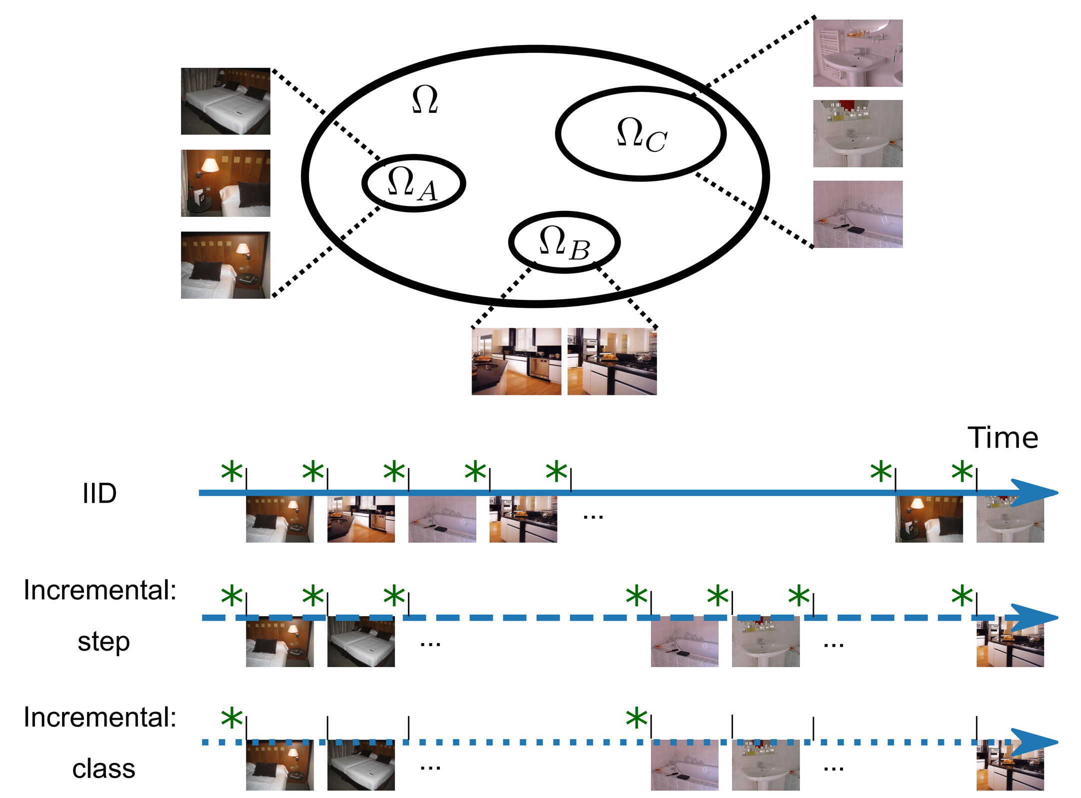
General Context: Big datasets and continual learning systems demand online algorithms. In addition, the input data of most embedded systems are not independent and identically distributed (iid), but correlated and non-balanced [1].

Problem: Expectation Maximization is a standard probabilistic method frequently used in the case of latent variables. It has been adapted for online learning with the introduction of a stochastic integration step on the expectation part. However, this algorith is slow and, most importantly, the guarantees of convergence are only valid in the case of iid samples [2]

Solution: We propose to constrain the online Expectation-Maximization on the Fisher distance between the reference parameters and the current update. The reference parameters are updated either at each iteration or by the end of a class [3].

Experiment: We tested our proposal with model data from PPCA, considering iid samples, incremental samples and, in the last one, two protocols to update reference parameters: step wise and class wise. We evaluated our proposal in terms of convergence, consolidation and interference [4].

Experimental Protocol



Formalism

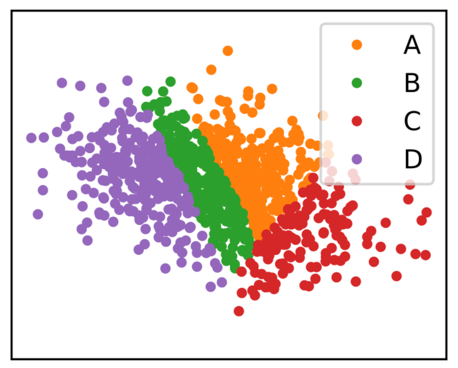
Batch	E-step	$Q(\theta, \theta_k) = \mathbb{E}_{Z\|X,\theta_k}\left[\mathcal{L}(X, Z; \theta_k)\right]$
	M-step	$\theta_{k+1} = \arg\max_{\theta} Q(\theta, \theta_k)$
Online	Stochastic	$S^i_{k+1} = S^i_k + \gamma\left(s^i - S^i_k\right)$
Nat-oEM	Regularized	$\theta_{k+1} = \arg\max_{\theta}\left(Q(\theta_k, \theta) - \beta\|\theta - \theta^*\|_{F^*}\right)$

Nat-oEM PPCA

Model:

$$X = WZ + \mu + \varepsilon$$
$$Z \sim \mathcal{N}(0, \mathbb{I})$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2\mathbb{I})$$

Data:



E-step: Sufficient statistics

$$s^0(x, \mu) = (x - \mu)^T(x - \mu)$$
$$s^1(x, z) = (x - \mu)z^T$$
$$s^2(x, \sigma, M) = \sigma^2 M^{-1} + zz^T$$
$$s^3(x) = x$$

E-step: Stochastic Integration

$$S^i_{k+1} = S^i_k + \gamma\left(s^i - S^i_k\right)$$

M-step:

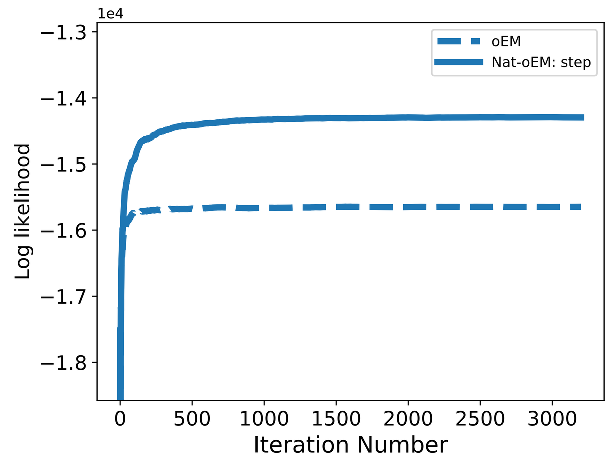
$$\mu_{k+1} = S^3$$
$$W_{k+1} = S^1 {S^2}^{-1}$$
$$\sigma^2_{k+1} = \frac{1}{d}(S^0 - 2\,\mathrm{Tr}(S^1 W^T) + \mathrm{Tr}(S^2 W^T W))$$

R-step:

$$\mu_{k+1\_reg} = \mu_{k+1} - \beta {F^*_\mu}^{-1}(\mu_{k+1} - \mu^*)$$
$$W_{k+1\_reg} = W_{k+1} - \beta {F^*_W}^{-1}(W_{k+1} - W^*)$$
$$\sigma^2_{k+1\_reg} = \sigma^2_{k+1} - \beta {F^*_{\sigma^2}}^{-1}\left(\sigma^2_{k+1} - {\sigma^2}^*\right)$$

Iid: Convergence

After few iterations, the log-likelihood of Nat-oEM was larger than the oEM. Furthermore it remains parallel over time

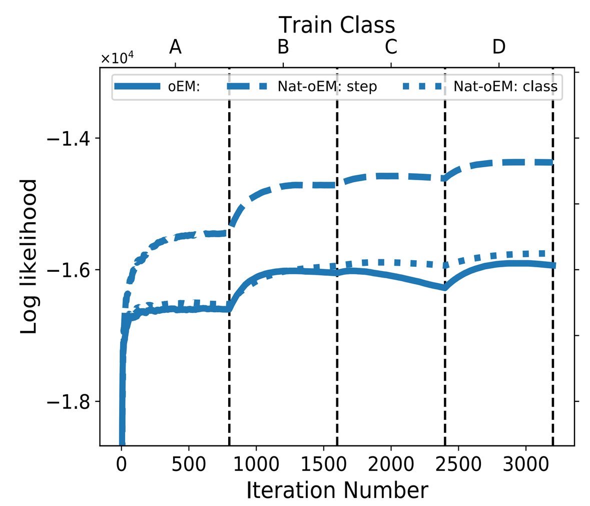


Incremental: Convergence

Without any constraint the overall likelihood decreases over time.

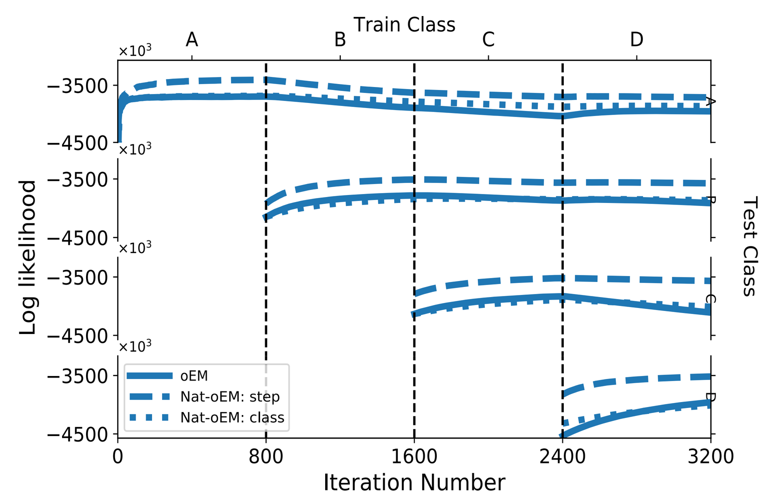
This outcome is avoided with Nat-oEM: class.

With Nat-oEM: step, the likelihood is boosted and monotonically increasing.



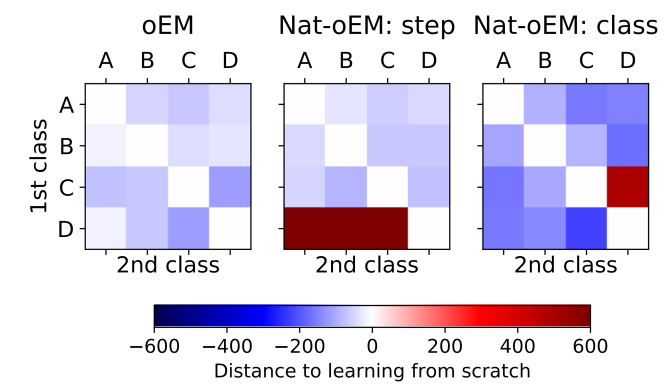
Incremental: Consolidation

Similarly, the class wise loglikelihood shows that there is forgetting with oEM, this forgetting is avoided with Nat-oEM: class and the likelihood is boosted with Nat-oEM: step.



Incremental: Interference

There is a slight negative interference with oEM. This tendence is similar wtih Nat-oEM step, except class D that has strong positive interference. Nat-oEM class induces strong negative interference except between class C and D.



Conclusions

- Nat-oEM converges faster than oEM.
- Nat-oEM avoids forgetting while oEM does not.
- Overall, Nat-oEM: step performance is better than Nat-oEM class, but increases the computational load.
- Nat-oEM: step introduces positive interference, while Nat-oEM: class introduces negative.

Bibliography & Acknowl.

[1] A. Gepperth and C. Karaoguz, “A Bio-Inspired Incremental Learning Architecture for Applied Perceptual Problems”, 2016.
[2] O. Cappé and E. Moulines, “Online EM Algorithm for Latent DataModels”, 2007.
[3] S.-i. Amari, “Natural Gradient Works Efficiently in Learning”, 1998.
[4] N. Díaz-Rodríguez et al. “Don’t forget, there is more than forgetting: new metrics for continual learning”, 2018.

This research has been made under financial support from the French Government Defense Procurement Agency (DGA)