Tutorial on Approximate Sorting
(SLT 2021)

Paolo Penna

May 31, 2021

These are **informal** notes whose purpose is to introduce a new problem (approximate sorting - more details in the **original paper** [BCB13]). The corresponding **exercises** are reported below and they will be discussed during the exercise class (31 May).

1 Setting (noisy comparisons ⇒ approximate sorting)

We have N objects

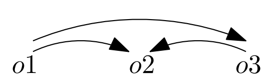
$$o_1 \ldots, o_i, \ldots, o_N$$

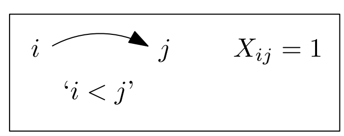
which we would like to sort (rank) based on pairwise **noisy comparisons** represented by a data set X where

$$X_{ij} = \begin{cases} 1 & \text{if `}o_i < o_j\text{'} \\ 0 & \text{if `}o_i > o_j\text{'} \end{cases} .$$

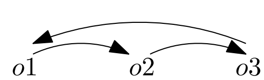
(Note: $X_{ii} = 0$ and $X_{ij} = 1 - X_{ji}$ for all i and $j \neq i$.)

Example 1. *For three objects the comparison results may look like this:*





or maybe like this



The feasible solutions are all **permutations** c over the items, where $c_i = \ell$ denotes the position (rank) of object o_i. Given the observed data (pairwise comparisons) we assign higher cost to solutions which **violate 'many'** comparisons in X.

Cost function = number of violated comparisons

$$\mathcal{R}^{sort}(c|X) := |\{(i,j) \mid X_{ij} = 1 \text{ but } c_i > c_i\}| \tag{1}$$

2 Mean Field Approximation

We try to follow an approach similar to the problems in Lecture 8

1. A solution is a matching (assignment) of each item to one position (rank).
2. Items choose their own rank **independently** of each other.

A binary matrix M can be used to represent our solutions (ordering)

$$M_{i\ell} = \begin{cases} 1 & \text{if } o_i \text{ gets rank } \ell \\ 0 & \text{if `}o_i > o_j\text{'} \end{cases} \qquad \text{(assignment of items to ranks)} \tag{2}$$

subject to

$$\sum_{\ell} M_{i\ell} = 1 \qquad \text{(each item gets a unique rank)} \tag{3}$$

$$\sum_{i} M_{i\ell} = 1 \qquad \text{(each rank gets a unique item)} \tag{4}$$

Mean field formulation

We assign weights (costs) to each possible item-rank assignment '$i \to \ell$' which defines our probability distribution Q:

$$\mathcal{E}_{i\ell} \quad \Rightarrow \quad q_{i\ell} := \frac{e^{-\beta \mathcal{E}_{i\ell}}}{\sum_h e^{-\beta \mathcal{E}_{ih}}}$$

$$\Rightarrow \quad Q(M|\mathcal{E}) = q_{1M(1)} \cdot q_{2M(2)} \cdots q_{NM(N)}$$

where

$$M(i) = \ell \Leftrightarrow M_{i\ell} = 1 \; .$$

Exercise 1. *Discuss whether our distribution Q always produces an M which corresponds to a feasible solution (rank).*

Exercise 2. *For any M which corresponds to a feasible solution (i.e. (3)-(4) hold) rewrite the original cost function (1) as $\mathcal{R}^{sort}(M|X)$.*

Remark 1. *The mean field formulation above is essentially replacing the original cost function $\mathcal{R}^{sort}(\cdot|X)$ by*

$$\mathcal{R}^{MF}(M|\mathcal{E}) := \sum_i \sum_{\ell} M_{i\ell}\mathcal{E}_{i\ell} \; . \tag{5}$$

Recall that, among all $\mathcal{E}$ we want the 'closest' to the original cost function, that is the one minimizing the Gibbs free energy.

To solve this mean field formulation we can reapply part of the formulation for clustering (in fact mean field for 'general' assignment problems – Lecture 8):

Mean Field equations:

$$q_{i\ell} = \frac{e^{-\beta \mathcal{E}_{i\ell}}}{\sum_h e^{-\beta \mathcal{E}_{ih}}} \tag{6}$$

$$\mathcal{E}_{i\ell} = \mathbb{E}_{M \sim Q_{i\to\ell}}[\mathcal{R}^{sort}(M|X)] \tag{7}$$

where in (7) we choose M according to Q restricted to assigning object o_i to rank ℓ.

Exercise 3. *Show that* (7) *implies*

$$\mathcal{E}_{i\ell} = \sum_j \left(\sum_{\underline{\ell}<\ell} X_{ij} q_{j\underline{\ell}} + \sum_{\bar{\ell}>\ell} X_{ji} q_{j\bar{\ell}} \right) + constant \tag{8}$$

Hint: *Linearity of expectation + look at the probability of a fixed comparison to be violated.*

Where do we use the assumption '$Q_{i\to\ell}$' in (7)*?*

Where does the 'constant' term in (8) *come from?*

Why we can use the EM algorithm to solve the mean field equations (6)-(7)*?*

Exercise 4. *A generic M may violate both* (3)-(4) *and thus may not be a valid rank. We propose to replace the cost function $\mathcal{R}^{MF}$ by the following one:*

$$\mathcal{R}^{MF}_{fix}(M|\mathcal{E}) := \sum_i \sum_\ell M_{i\ell}\mathcal{E}_{i\ell} + \sum_\ell \lambda_\ell \left(\sum_j M_{j\ell} - 1 \right) \ . \tag{9}$$

Generalize (6)-(7)-(8) *to this mean field cost function $\mathcal{R}^{MF}_{fix}$.* ***Hint:*** *Look at smooth k-means in Lecture 8.*

How are the constraints (3)-(4) *incorporated into* (9)*?*

Compare your generalization of solution for the mean field equations (6)-(7)-(8) *with the those in* ***original paper*** *[BCB13]*

3 Posterior Agreement

Two data sets:

$$X' \qquad R'(c) := \mathcal{R}^{sort}(c|X') \quad p'_\beta(c) := p_\beta(c|X') = \frac{e^{-\beta R'(c)}}{Z'_\beta} \qquad Z'_\beta = \sum_c e^{-\beta R'(c)}$$

$$X'' \qquad R''(c) := \mathcal{R}^{sort}(c|X'') \quad p''_\beta(c) := p_\beta(c|X'') = \frac{e^{-\beta R''(c)}}{Z'_\beta} \qquad Z''_\beta = \sum_c e^{-\beta R''(c)}$$

and their joint cost

$$X' + X'' \quad R^+(c) := R'(c) + R''(c) \qquad p^+_\beta(c) = \frac{e^{-\beta R^+(c)}}{Z^+_\beta} \qquad Z^+_\beta = \sum_c e^{-\beta R^+(c)}$$

Recall that we are interested in a particular $\beta^* = \beta^*(X', X'')$:

$$\beta^* = \arg\max_\beta \; k_\beta(X, X'') \qquad k_\beta(X', X'') := \underbrace{\sum_c p_\beta(c|X')p_\beta(c|X'')}_{kernel\ PA} \tag{10}$$

Exercise 5. *Show that*

$$\sum_c p_\beta(c|X')p_\beta(c|X'') = \frac{Z^+_\beta}{Z'_\beta \cdot Z''_\beta} \tag{11}$$

Factorizable (mean field) distributions ⇒ compute posterior agreement

1. Rewrite the posterior agrement in terms of partition functions (11);
2. Use mean field approximation where the partition function can be computed.

For the second step, do the following:

$$p'_\beta(\cdot) \qquad \Rightarrow \qquad \tilde{p}'_\beta(\cdot) = Q(\cdot|\mathcal{E}') \tag{12}$$

$$p''_\beta(\cdot) \qquad \Rightarrow \qquad \tilde{p}''_\beta(\cdot) = Q(\cdot|\mathcal{E}'') \tag{13}$$

Exercise 6. *Based on the mean field approximation, suggest a procedure which finds an 'approximate/reasonable' $\tilde{\beta}^*$ for the posterior agreement (11). Specifically, suppose we have discretized the possible values for β, that is, we want some*

$$\tilde{\beta}^* \in \{\beta_1, \beta_2, \ldots, \beta_d\} \; .$$

Your procedure should be (computationally) efficient, in particular should not rely on computing the partition functions Z'_β, Z''_β and Z^+_β.

How does your procedure work for $d = 2$?

Does the mean field part depend on $\beta_1, \beta_2, \ldots$?

Discuss how the experiments in the ***original paper*** *[BCB13] can be performed (Fig 3(a)).*

References

[BCB13] Ludwig Busse, Morteza Haghir Chehreghani, and Joachim M Buhmann. Approximate sorting. In *German Conference on Pattern Recognition*, pages 142–152. Springer, 2013.

A Solutions (SLT 21 Series 11 – Approximate Sorting)

This part contains the solutions to the exercises.
It is published as both a standalone file and as the appendix of the tutorial.
Hyperlinks work olnly on the tutorial version.

Solution to Exercise 1. No it does not and in particular (4) can be violated. The constraint (3) is guaranteed by definition, since $q_{i\ell}$ is normalized

$$q_{i\ell} := \frac{e^{-\beta\mathcal{E}_{i\ell}}}{\sum_h e^{-\beta\mathcal{E}_{ih}}}$$

and it is the probability that $i \mapsto \ell$ (a unique $M_{i\ell}$ is equal to 1.) □

Solution to Exercise 2.

$$\mathcal{R}^{sort}(M|X) = \sum_i \sum_j \sum_\ell \sum_{\bar{\ell}>\ell} X_{ij} M_{i\bar{\ell}} M_{j\ell}$$

□

Solution to Exercise 3. Rewrite the expected cost by looking at the contribution of each violated comparison:

$$\underbrace{X_{ij} = 1}_{i<j} \Rightarrow i \to \ell \text{ and } j \to \underline{\ell} \text{ with } \underline{\ell} < \ell \tag{14}$$

$$\underbrace{X_{ji} = 1}_{j<i} \Rightarrow i \to \ell \text{ and } j \to \underline{\ell} \text{ with } \bar{\ell} > \ell \tag{15}$$

These are only the comparisons involving i. The others, not involving i, are

$$\underbrace{X_{ab} = 1}_{a<b} \Rightarrow a \to \bar{\ell} \text{ and } b \to \underline{\ell} \text{ with } \underline{\ell} < \bar{\ell} \tag{16}$$

Putting all these contributions together into the expectation

$$\mathcal{E}_{i\ell} = \mathbb{E}_{M\sim Q_{i\to\ell}}[\mathcal{R}^{sort}(M|X)] = \underbrace{\sum_j \sum_{\underline{\ell}<\ell} X_{ij} q_{j\underline{\ell}} + \sum_j \sum_{\bar{\ell}<\ell} X_{ji} q_{j\bar{\ell}}}_{:=\Sigma_{i\ell}} + \underbrace{\sum_{a\neq i} \sum_{b\neq i} X_{ab} \sum_{\underline{\ell}} \sum_{\bar{\ell}>\underline{\ell}} q_{a\bar{\ell}} q_{b\underline{\ell}}}_{constant_i} \tag{17}$$

where $constant_i$ **does not depend on** ℓ. Note that this implies that $q_{i\ell}$ in (6) does not depend on $constant_i$ but only on the $\Sigma_{i\ell}$ term:

$$q_{i\ell} = \frac{e^{-\beta(\Sigma_{i\ell}+constant_i)}}{\sum_h e^{-\beta(\Sigma_{ih}+constant_i)}} = \frac{e^{-\beta\Sigma_{i\ell}}}{\sum_h e^{-\beta\Sigma_{ih}}} \tag{18}$$

Since the term $\Sigma_{i\ell}$ does not contain $q_{i\ell}$ we can apply the EM-algorithm. □

Solution to Exercise 4. In short, the function $\mathcal{R}^{MF}_{fix}$ has the effect of shifting the fields

$$\mathcal{E}_{i\ell} \to \hat{\mathcal{E}}_{i\ell} = \mathcal{E}_{i\ell} + \lambda_\ell$$

that is

$$Q(M|\mathcal{E}, \lambda) = Q(M|\hat{\mathcal{E}})$$

So, we can just reapply equations (6)-(7)-(8) with the shifted fields $\hat{\mathcal{E}}$ and get

First attempt (wrong):

$$q_{i\ell} = \frac{e^{-\beta(\mathcal{E}_{i\ell}+\lambda_\ell)}}{\sum_h e^{-\beta(\mathcal{E}_{ih}+\lambda_h)}} \tag{19}$$

$$\mathcal{E}_{i\ell} + \lambda_\ell = \mathbb{E}_{M\sim Q_{i\to\ell}}[\mathcal{R}^{sort}(M|X)] \tag{20}$$

and specifically

$$\mathcal{E}_{i\ell} + \lambda_\ell = \sum_j \left(\sum_{\underline{\ell}<\ell} X_{ij} q_{j\underline{\ell}} + \sum_{\overline{\ell}>\ell} X_{ji} q_{j\overline{\ell}} \right) + constant \tag{21}$$

My formula is **not** the same as in the **original paper** [BCB13], which does **not contain the 'λ_ℓ' term**, and thus each $\mathcal{E}_{i\ell}$ can be computed directly from the q_{ih}'s:

$$\mathcal{E}_{ik} = \sum_{j\neq i} \left(\sum_{k'=1}^{k+1} X_{ij} q_{jk'} + \sum_{k'=k+1}^{N} X_{ji} q_{jk'} \right) + constant \tag{22}$$

I made a **'conceptual' mistake**, ignoring the fact that the mean field solution M may violate constraint (4) and that the coefficients λ_ℓ translate this constraint into a 'soft constraint'

$$\sum_\ell \lambda_\ell \left(\sum_j M_{j\ell} - 1 \right) \ .$$

The correct Equation (22) can be obtained by similar strategy described in Exercise 8 (Problem 2):

1. Add the soft constraint (Lagrange multipliers) to the original cost function:

$$\mathcal{R}^{sort}_{fix}(M|X) := \mathcal{R}^{sort}(M|X) + \sum_\ell \lambda_\ell \left(\sum_j M_{j\ell} - 1 \right) \tag{23}$$

2. This means that the soft constraint is added to the usual Gibbs free energy $G(\cdot)$:

$$G_{fix}(Q) = G(Q) + \sum_\ell \lambda_\ell \left(\sum_j q_{j\ell} - 1 \right) \ .$$

3. When computing the derivatives $\frac{\partial}{\partial q_{i\ell}} G_{fix}(Q)$ with respect to $q_{i\ell}$ (extremality conditions) the soft constraint introduces a 'λ_ℓ' term as follows:

$$E_{M \sim Q_{i \mapsto \ell}}[\mathcal{R}^{sort}(M|X)] + \lambda_\ell \stackrel{!}{=} \frac{1 + \log q_{i\ell}}{\beta} \tag{24}$$

This leads to the correct equations:

Mean Field equations (correct version):

$$q_{i\ell} = \frac{e^{-\beta(\mathcal{E}_{i\ell} + \lambda_\ell)}}{\sum_h e^{-\beta(\mathcal{E}_{ih} + \lambda_h)}} \tag{25}$$

$$\mathcal{E}_{i\ell} = \sum_j \left(\sum_{\underline{\ell} < \ell} X_{ij} q_{j\underline{\ell}} + \sum_{\overline{\ell} > \ell} X_{ji} q_{j\overline{\ell}} \right) + constant \tag{26}$$

Remark 2 (typo in original paper)**.** *Note that the formula in the original paper contains a typo: '$k+1$ in the first summation should be '$k-1$'.*

Remark 3 (lagrangian multipliers λ_ℓ's)**.** *The purpose of the coefficients λ_ℓ's is to enforce the condition (4). The latter can be rewritten as $e^{\beta\lambda_\ell} = e^{\beta\lambda_\ell} \sum_i q_{i\ell}$ which typically leads to a method to estimate the right λ_ℓ's for given $q_{i\ell}$'s (other equalities are also possible).*

Remark 4. *Note that constraint (3) is always satisfied because of (25).*

□

Solution to Exercise 5. The identity is true for any problem, that is, for any cost function $R(c|X)$ (in our problem we take $R = \mathcal{R}^{sort}$):

$$\sum_c p_\beta(c|X')p_\beta(c|X'') = \sum_c \frac{e^{-\beta R(c|X')}}{Z'_\beta} \frac{e^{-\beta R(c|X'')}}{Z''_\beta} = \sum_c \frac{e^{-\beta[R(c|X') + R(c|X'')]}}{Z'_\beta \cdot Z''_\beta} \tag{27}$$

$$= \sum_c \frac{e^{-\beta R^+(c)}}{Z'_\beta \cdot Z''_\beta} \tag{28}$$

$$= \frac{Z^+_\beta}{Z'_\beta \cdot Z''_\beta} \tag{29}$$

□

Solution to Exercise 6. For a fixed $\dot{\beta}$ we approximate the two Gibbs distributions as illustrated in (12)-(13). Because the two distributions factorize, we can calculate the corresponding partition functions. From Exercise 5 we can compute the corresponding kernel PA (the partition function corresponding to Z^+_β also comes from a mean field approximation in the same way). Note the following:

1. The mean fields $\mathcal{E}'$ and $\mathcal{E}''$ do depend on $\dot{\beta}$.

2. If we had to compare two possible values for β, say β_1 and β_2, we simply apply the procedure above and compare their resulting (mean field/approximate) kernel PA's.

3. We could thus discretize the possible values (essentially choose some d) and find the best $\tilde{\beta}^*$ by generalizing the previous step.

The mean field should use the 'enhanced' version in Exercise 4 with coefficients λ_ℓ's (which also depend on $\dot{\beta}$). $\square$