**Expectation-Maximization for Learning Determinantal Point Processes**

Jennifer Gillenwater
jing@cis.upenn.edu

Alex Kulesza
kulesza@umich.edu

Emily Fox
ebfox@cs.washington.edu

Ben Taskar
taskar@cs.washington.edu

---

**TASK: SUBSET SELECTION**

Example: product recommendation

Two goals: relevance and diversity.

---

**SIMILARITY KERNEL**

$K = \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ 0.38 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix}$

---

**PARAMETRIC KERNEL**

K assumed Gaussian, poly etc.

Affandi et al. (ICML 2014)

---

**DETERMINANTAL POINT PROCESS**

$Y \sim DPP(K) \implies P(Y \subseteq Y) = \det(K_Y)$

---

**BASELINE: PROJECTED GRADIENT**

Gradient: $\frac{\partial \log L(K)}{\partial K} \propto \sum_{i=1}^{T} (K - I_T)^{-1}$

---

**GENERATIVE DPP SAMPLING**

Eigendecompose $K = |\Lambda| V^T$

Sample hidden eigenvectors $J$

---

**GENERATIVE DPP SAMPLING**

$Y \sim DPP(K)$

---

**EXPECTATION-MAXIMIZATION**

Main idea: Exploit hidden variable $J$ to develop EM-style optimization.

---

**PRIOR WORK**

Quality learning

Learn weight for each row of $K$ (Kulesza and Taskar (ICML 2011))

---

**IMPLEMENTATION**

DPP max likelihood learning is (conjectured) NP-hard

---

**NP-METHOD**

---

**LOG-LIKELIHOOD IS NON-CONCAVE**

Training data: $\bullet \bullet \bullet$

---

**LOG-LIKELIHOOD RESULTS**

---

**PRODUCT RECOMMENDATION DATASET**

---

**SUM-OF-EXPERTS MODEL**

Weighted sum of fixed DPPs (Kulesza and Taskar (ICML 2011))

---

**RELEVANCE SAMPLING**

Marginalization:

$DPP(Y) \propto \prod_{i=1}^{N} \prod_{j \in Y} \log(p_{ij})$