

# P4-5 ブレグマン距離を用いた 密度比推定の統一的枠組み

杉山将(東工大)・鈴木大慈(東大)・金森敬文(名大)

$$\text{密度比: } r(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$$

- 密度比は様々なデータ解析に有用:
  - 重点サンプリング (共変量シフト適応, ドメイン適応, マルチタスク学習)
  - 二標本問題 (二標本検定, 外れ値検出, 変化点検知)
  - 相互情報量推定 (独立性検定, 特徴選択・抽出, 独立成分分析, 因果推定)
  - 条件付き確率推定 (可視化, 状態遷移推定, 確率的パターン認識)
- $p(\boldsymbol{x}), q(\boldsymbol{x})$  を推定せず,  $r(\boldsymbol{x})$  を直接推定して精度を向上:
  - 確率的分類法 (Qin, Biometrika1998)
  - 積率適合法 (Huang *et al.*, NIPS2006)
  - 密度適合法 (Sugiyama *et al.*, NIPS2007, AISM2008)
  - 密度比適合法 (Kanamori *et al.*, NIPS2008, JMLR2009)

## ■ 密度比推定の統一的枠組み:

- **ブレグマン距離**のもとで, 密度比のモデル  $\hat{r}(\boldsymbol{x})$  を真の密度比  $r(\boldsymbol{x})$  に適合させる

$$\min_{\hat{r}} \left[ \int q(\boldsymbol{x}) \left[ \nabla f(\hat{r}(\boldsymbol{x})) \hat{r}(\boldsymbol{x}) - f(\hat{r}(\boldsymbol{x})) \right] d\boldsymbol{x} - \int p(\boldsymbol{x}) \nabla f(\hat{r}(\boldsymbol{x})) d\boldsymbol{x} \right]$$

- ブレグマン距離はユークリッド距離の一般化
- 経験近似が即座に得られる  $\Rightarrow$  密度推定不要

## ■ 既存の密度比推定法が全てこの枠組みに含まれる!

- **確率的分類法**:  $f(t) = t \log t - (1+t) \log(1+t)$
- **密度適合法**:  $f(t) = t \log t - t$
- **積率適合法, 密度比適合法**:  $f(t) = (t-1)^2/2$

## ■ 新たな密度比推定法:

- **べき距離** (Basu *et al.*, Biometrika 1998):  $f(t) = \alpha^{-1} (t^{1+\alpha} - t)$
- モデルからの外れ値に対して**ロバスト**  $\alpha > 0$