

D-49 近似ベイズ推論としての dropoutとその最適化

前田新一 京都大学

DeepFace (CVPR, 2014)

顔画像 → 4030人の個人識別

パラメータ数**1.2億個** >> 学習データ数**400万個**

← 過学習しない?

dropoutは**近似ベイズ推論**?

既存のdropoutの解釈

$$\begin{aligned}\log p(D | \theta) &= \log \left(\sum_{\mathbf{z}} \left(\prod_{t=1}^T p(y_t | \mathbf{x}_t, \mathbf{z}, \theta) \right) p(\mathbf{z}) \right) \\ &\geq \sum_t \sum_{\mathbf{z}} p(\mathbf{z}) \log p(y_t | \mathbf{x}_t, \mathbf{z}, \theta)\end{aligned}$$

提案するdropoutの解釈

$$\begin{aligned}\log p(D | \theta) &= \log \left(\sum_{\mathbf{z}} \left(\prod_{t=1}^T p(y_t | \mathbf{x}_t, \mathbf{z}, \theta) \right) p(\mathbf{z}) \right) \\ &\geq \sum_t \sum_{\mathbf{z}} q(\mathbf{z}) \log p(y_t | \mathbf{x}_t, \mathbf{z}, \theta) + \sum_{\mathbf{z}} q(\mathbf{z}) \log p(\mathbf{z}) - \sum_{\mathbf{z}} q(\mathbf{z}) \log q(\mathbf{z})\end{aligned}$$

既存のdropoutは、 $q(\mathbf{z})=p(\mathbf{z})$ に固定 ➡ $q(\mathbf{z})$ を最適化してみよう