BatchBALD

Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning

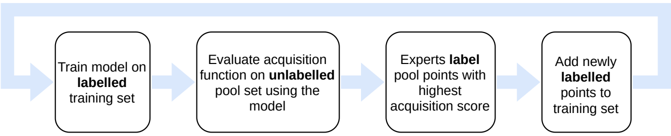
Andreas Kirsch*, Joost van Amersfoort*, Yarin Gal

OATML, Department of Computer Science, University of Oxford

{andreas.kirsch, joost.van.amersfoort, yarin}@cs.ox.ac.uk

Active Learning

A key problem in deep learning is **data efficiency**. In Active Learning, we iteratively acquire labels for only the **most informative** data points.



BALD Acquisition Function[1]

We implement a Bayesian Neural Network using dropout VI[2] and define the acquisition function a as follows:

$$a_{\mathrm{BALD}}\left(\{x_1, \ldots, x_b\}, \mathrm{p}(\boldsymbol{\omega} \mid \mathcal{D}_{\mathrm{train}})\right) := \sum_{i=1}^{0} \mathbb{I}\left(y_i ; \boldsymbol{\omega} \mid x_i, \mathcal{D}_{\mathrm{train}}\right)$$

$$\mathbb{I}\left(y ; \boldsymbol{\omega} \mid x, \mathcal{D}_{\mathrm{train}}\right) = \mathbb{H}\left(y \mid x, \mathcal{D}_{\mathrm{train}}\right) - \mathbb{E}_{\mathrm{p}(\omega \mid \mathcal{D}_{\mathrm{train}})}\left[\mathbb{H}\left(y \mid x, \boldsymbol{\omega}, \mathcal{D}_{\mathrm{train}}\right)\right]$$

First term captures general uncertainty of model.
Second term captures the uncertainty of a given draw of the model parameters

Score is high when model is uncertain in general (high entropy), but per parameter sample certain (expectation of sample entropy low).

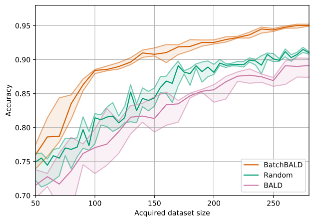
Batch Acquisitions

In practice, we acquire **the top-b highest scoring points:**

$$\{x_1^*, \ldots, x_b^*\} = \underset{\{x_1, \ldots, x_b\} \subseteq \mathcal{D}_{\mathrm{pool}}}{\arg\max} \; a\left(\{x_1, \ldots, x_b\}, \mathrm{p}(\boldsymbol{\omega} \mid \mathcal{D}_{\mathrm{train}})\right)$$

But naively applying BALD this way leads to redundant acquisitions, **under performing random acquisitions**!

Results on **Repeated MNIST:**



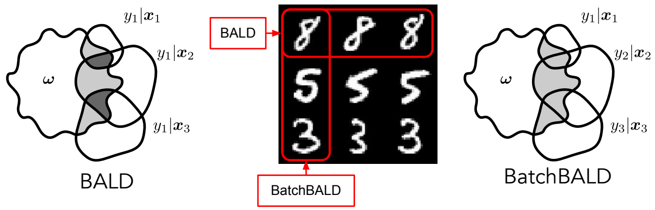
BatchBALD

We propose to compute BALD over a **batch** of points:

$$a_{\mathrm{BatchBALD}}\left(\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_b\}, \mathbf{p}(\boldsymbol{\omega} \mid \mathcal{D}_{\mathrm{train}})\right) = \mathbb{I}\left(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_b ; \boldsymbol{\omega} \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_b, \mathcal{D}_{\mathrm{train}}\right)$$

Expanding the Mutual Information:

$$\mathbb{I}\left(y_{1:b} ; \boldsymbol{\omega} \mid \boldsymbol{x}_{1:b}, \mathcal{D}_{\mathrm{train}}\right) = \mathbb{H}\left[\left(y_{1:b} \mid \boldsymbol{x}_{1:b}, \mathcal{D}_{\mathrm{train}}\right) - \mathbb{E}_{\mathbf{p}(\boldsymbol{\omega} \mid \mathcal{D}_{\mathrm{train}})} \mathbb{H}\left[\left(y_{1:b} \mid \boldsymbol{x}_{1:b}, \boldsymbol{\omega}, \mathcal{D}_{\mathrm{train}}\right)\right.\right.$$



BALD counts the dark areas double, while **BatchBALD** correctly computes the surface of the overlapping area

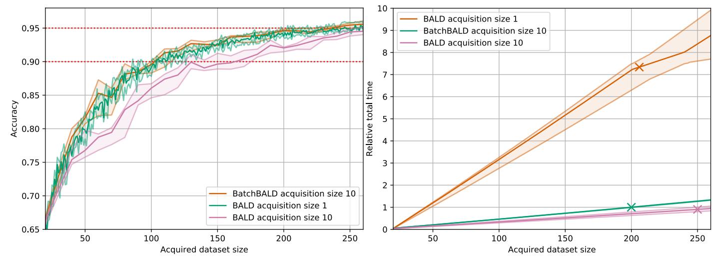
Computing BatchBALD

Computing **joint-entropy** exact requires evaluating exponential amount of candidates. In **BatchBALD**, we compute a **greedy approximation** and build up acquisition batch one by one. We show the approximation is **submodular** with an error bounded by $1 - \frac{1}{e}$.

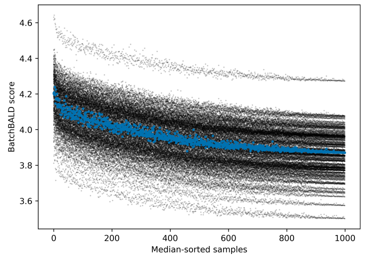
Definition: A function f defined on subsets of Ω is called **submodular** if for every set $A \subset \Omega$ and two non-identical points $x, y \in \Omega \backslash A$:

$$f(A \cup \{x, y\}) - f(A) \leq (f(A \cup \{x\}) - f(A)) + (f(A \cup \{y\}) - f(A))$$

MNIST

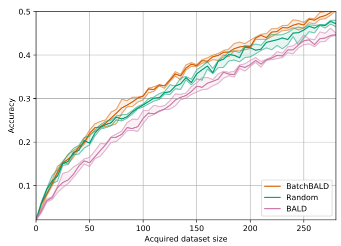


Consistent Dropout



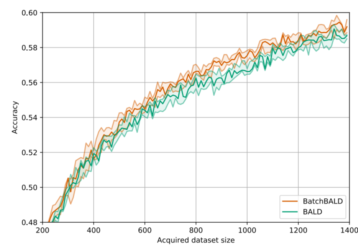
Computing BatchBALD requires keeping dropout masks constant when evaluating the acquisition score across the unlabelled pool set. As a side-effect it **reduces variance** when computing acquisition score! Also useful in BALD and other applications using dropout VI.

EMNIST



An extension of MNIST containing letters:

CINIC-10



CINIC-10 is a combination of CIFAR and ImageNet:

CIFAR-10:

CINIC-10:

[1] Houlsby, Neil, et al. "Bayesian active learning for classification and preference learning." *arXiv preprint arXiv:1112.5745* (2011).

[2] Gal, Yarin, Riashat Islam, and Zoubin Ghahramani. "Deep bayesian active learning with image data." *Proceedings of the 34th International Conference on Machine Learning-Volume 70.* JMLR. org, 2017.