DeepMind
Online Learning in Contextual Bandits using Gated Linear Networks

Eren Sezener*, Marcus Hutter*, David Budden, Jianan Wang, Joel Veness

Contextual bandits

Algorithm: The contextual bandit problem

1: **Input:** Unknown reward function h
2: **Input:** Action set $\mathcal{A}$
3: **for** each timestep t **do**
4: Observe context x_t.
5: Perform action $a_t \in \mathcal{A}$
6: Observe reward $r_t \sim h(x_t, a_t)$
7: **end for**
8: **Goal:** Maximize $\mathbb{E}[\sum_t r_t]$.

Applications:
- Hyperparameter optimization
- Personalized advertising
- Clinical trial design
- ...

The canonical frequentist policy:

$$a_t \leftarrow \arg\max_{a \in \mathcal{A}} \{\text{reward}(x_t, a) + \ C \ \cdot \text{novelty}(x_t, a)\}$$

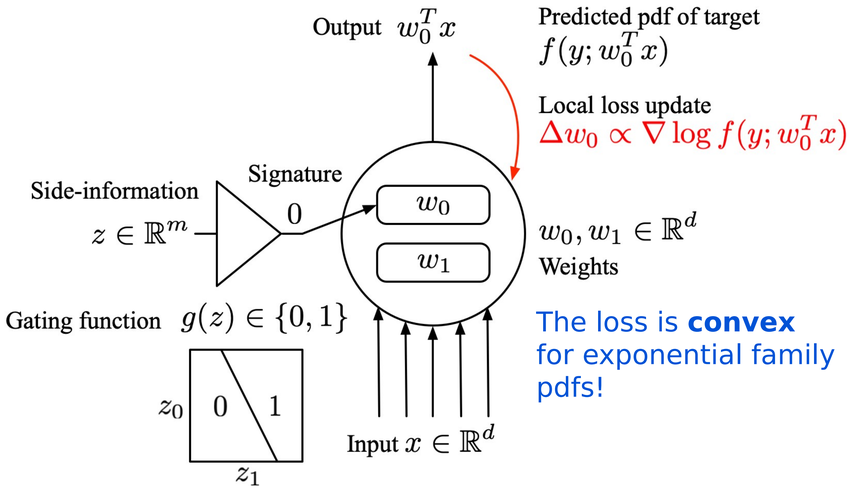
Use a Gated Linear Network

Obtain "for free"

Gated Linear Networks

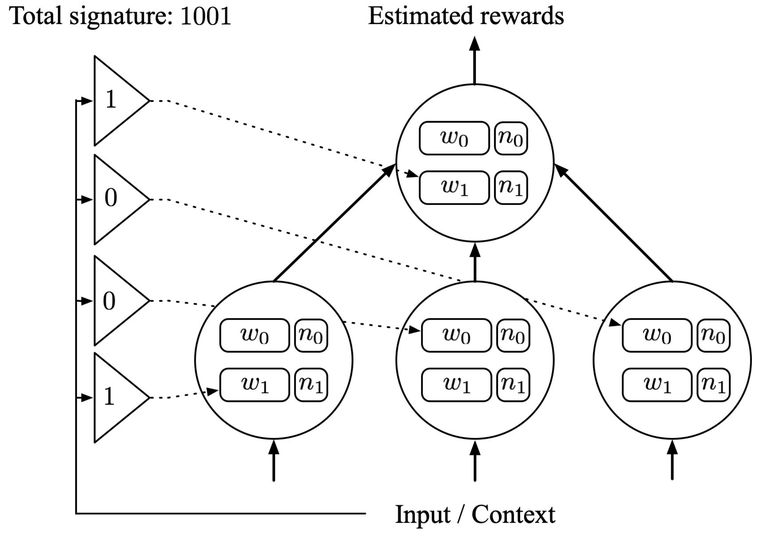
Gated Linear Networks [1, 2, 3] are a backpropagation-free family of deep neural networks with several desirable properties. GLNs are inherently:

- Fast, online learners
- Robust to forgetting
- Regret bounds
- Interpretable
- Distributional
- Compositional



The loss is **convex** for exponential family pdfs!

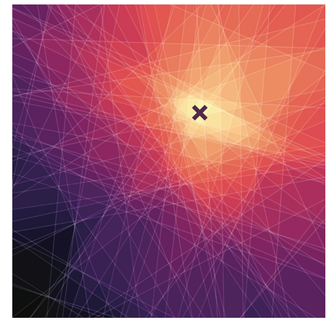
Gated Linear Contextual Bandits



Similar signatures imply similar data points.

Proposal: Each neuron keeps counts of encountered signatures.

low average count
⇒ a novel context



$$\text{novelty}(x_t, a) = \sqrt{\frac{\log t}{\text{softmin signature count of } (x_t, a)}}$$

Theorem: GLCB converges to the optimal policy almost surely.

Experiments

We compare against 9 neural "Bayesian" methods across 7 datasets provided by [1].

- All baselines are **offline**: storing the data and performing multiple passes.
- GLCB is **online**: one pass without storing any data.

Dataset	$\lvert\mathcal{D}\rvert$	$\lvert\mathcal{A}\rvert$	d	rewards
adult	45k	14	94	$\{0, 1\}$
census	2.5M	9	389	$\{0, 1\}$
covertype	581k	7	54	$\{0, 1\}$
statlog	43.5k	7	9	$\{0, 1\}$
financial	3.7k	8	21	$[0, 1]$
jester	19k	8	32	$[0, 1]$
wheel	-	5	2	$[0, 10]$

Algorithm	adult	census	covertype	statlog	financial	jester	wheel	mean rank
GLCB	**1**	**1**	5	**1**	2	4	2	**2.29**
BootRMS	2	2	**1**	3	4	**1**	8	3.00
Dropout	3	3	4	6	6	2	5	4.14
LinFullPost	5	8	6	5	**1**	6	**1**	4.57
NeuralLinear	7	5	7	2	3	7	3	4.86
RMS	4	4	3	7	5	3	9	5.00
BBB	6	7	2	4	8	5	6	5.43
ParamNoise	8	6	8	8	7	10	4	7.29
constSGD	9	9	9	9	9	8	6	8.43
BBAlphaDiv	10	10	10	10	10	9	10	9.86

Result: best mean rank across 7 datasets.

Regression problems are harder to learn in one pass?

References

[1] Veness, et al. "Online learning with gated linear networks" arXiv preprint arXiv:1712.01897 (2017)
[2] Veness, et al. "Gated linear networks" arXiv preprint arXiv:1910.01526 (2019)
[3] Riquelme, et al. "Deep Bayesian Bandits Showdown" ICLR (2018).

Join the GLN revolution at NeurIPS!

Spotlight Budden, David, et al. "Gaussian Gated Linear Networks" arXiv preprint arXiv:2006.05964 (2020)
Poster 17752 Wang, Jianan, et al. "A Combinatorial Perspective on Transfer Learning." arXiv preprint arXiv:2010.12268 (2020).
BeyondBackprop Workshop. Sezener, Eren, et al. "Gated Linear Networks and Extensions"

* *Equal contributions. All correspondence to esezener@google.com*