Feedback Coding for Active Learning

Gregory Canal, Matthieu Bloch, Christopher Rozell

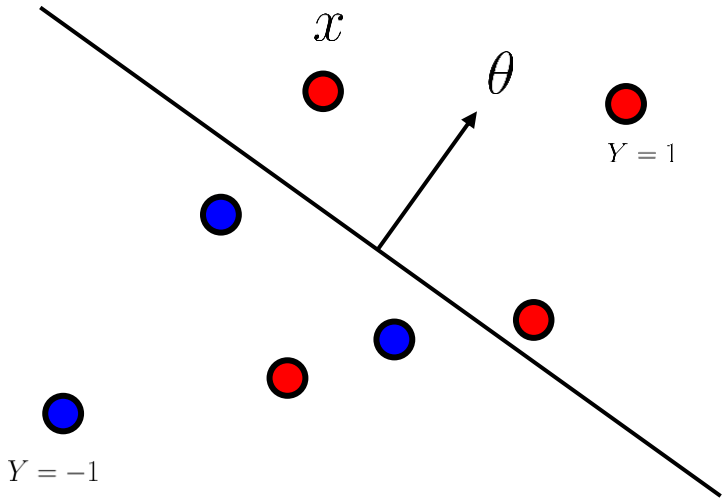
Georgia Institute of Technology

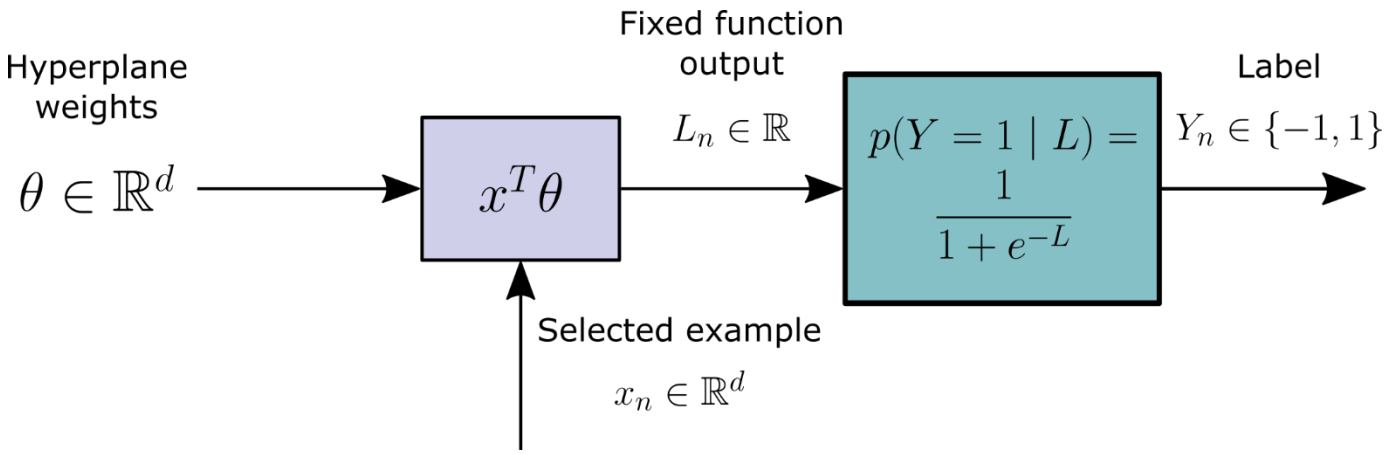
Active Learning as a Communications Model

$$p(Y = 1 \mid x, \theta) = \frac{1}{1 + e^{-x^T\theta}}$$

$$L = h_\theta(x) \coloneqq x^T\theta$$

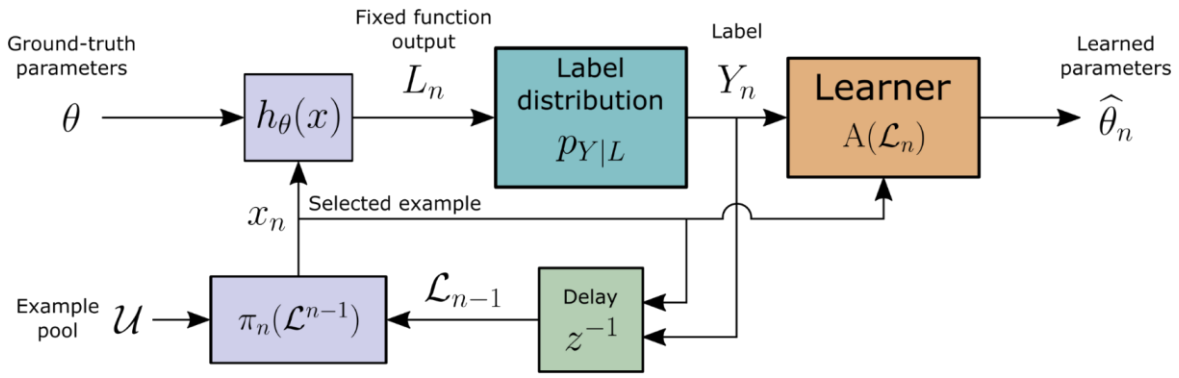
$$p(Y = 1 \mid x, \theta) = p(Y = 1 \mid L) = \frac{1}{1 + e^{-L}}$$



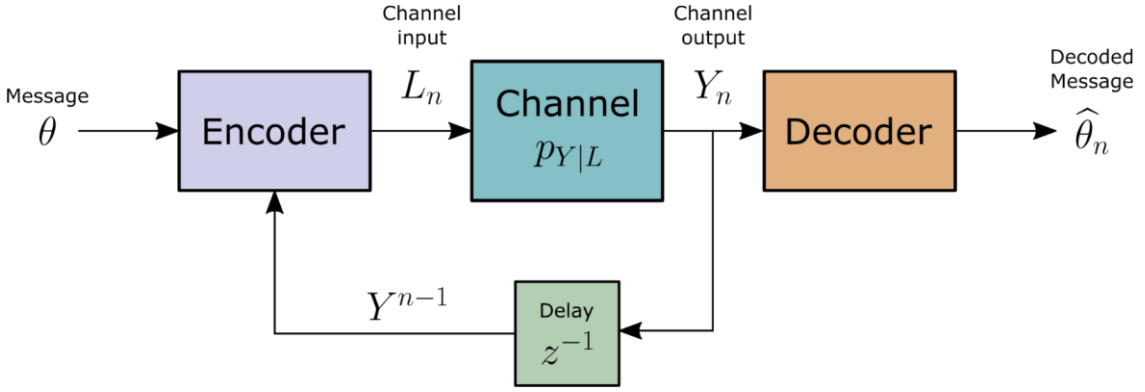


Active Learning as a Communications Model

Active Learning

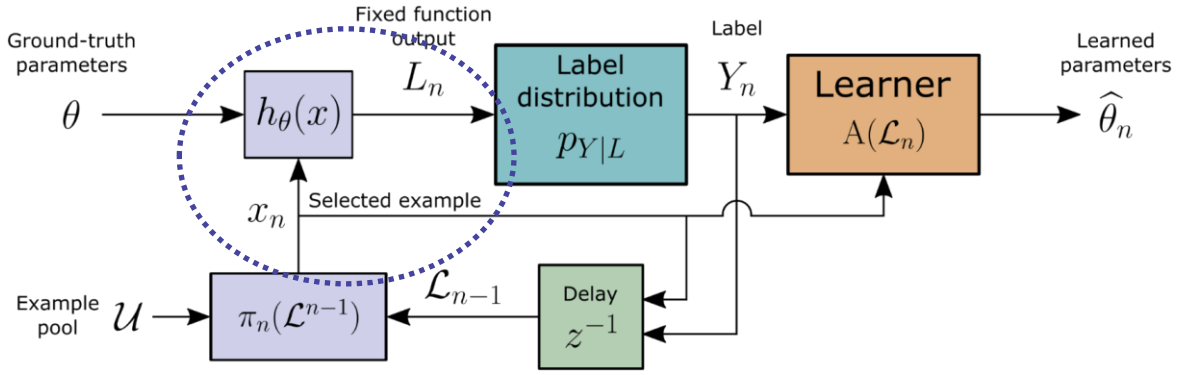


Feedback Coding

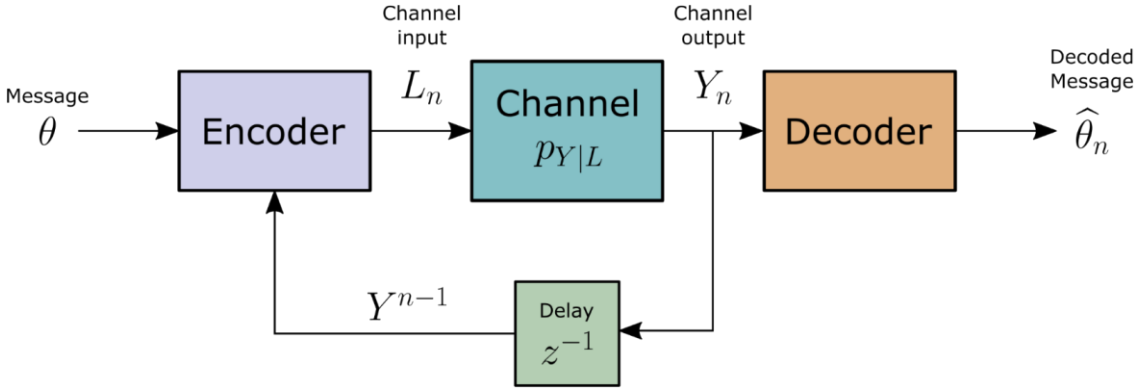


Active Learning as a Communications Model

Active Learning

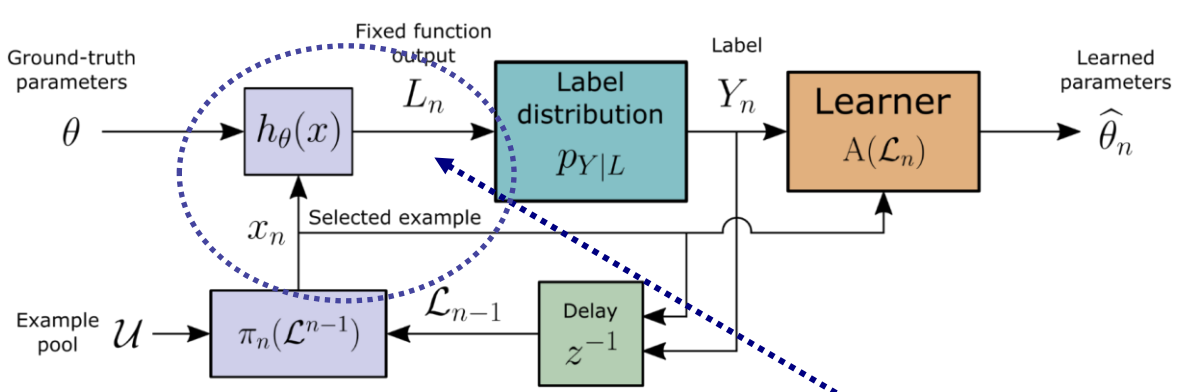


Feedback Coding

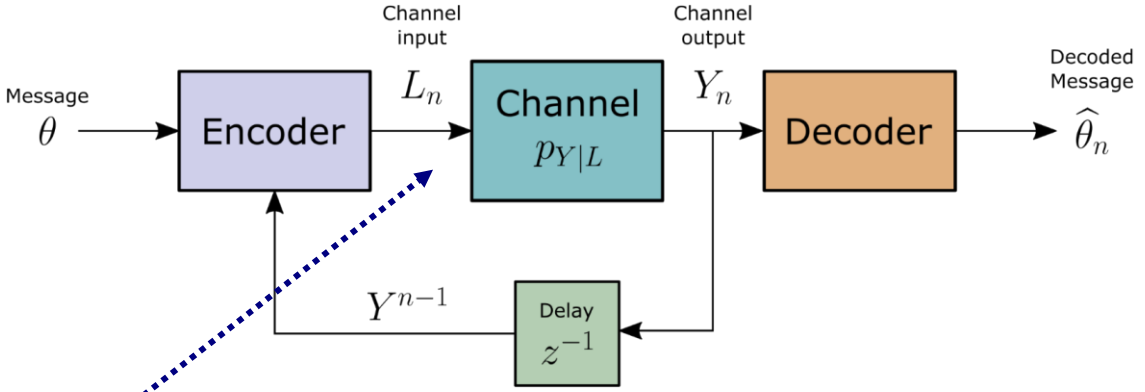


Active Learning as a Communications Model

Active Learning



Feedback Coding



Capacity-achieving distribution

$$p_L^* := \arg\max_{p_L} I(L;Y)$$

Result: BPSK in logistic regression

Approximate Posterior Matching (APM)

- Posterior Matching: design feedback encoder that induces capacity-achieving distribution on channel input (Ma and Coleman, 2011; Shayevitz and Feder, 2011)

- Not always possible to induce capacity-achieving distribution through example selection
  - **Result**: under mild assumptions, no such example exists in logistic regression

Approximate Posterior Matching (APM)

Approximate Posterior Matching (APM): select example such that channel input distribution is "as close as possible" to achieving capacity (as measured by Wasserstein distance)

$$x_n = \pi_n(\mathcal{L}_{n-1}) := \underset{x \in \mathcal{U}_n}{\arg\min} \; W_2(p_{L_n|\mathcal{L}_{n-1}}, p_L^*)$$

APM-LR

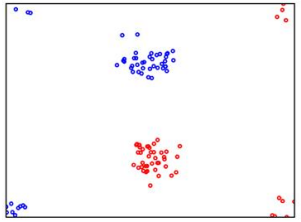
Approximate Posterior Matching for Logistic Regression (APM-LR):

$$p_{\theta|\mathcal{L}_{n-1}} \approx \mathcal{N}(\mu_n, \Sigma_n)$$

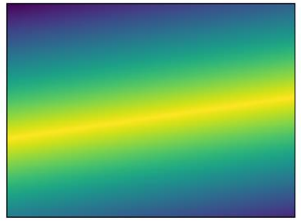
$$\pi_n(\mathcal{L}_{n-1}) = \arg\min_{x \in \mathcal{U}_n} \; (\mu_n^T x)^2 + \left(\sqrt{x^T \Sigma_n x} - \sqrt{\frac{2}{\pi} P} \right)^2$$

"Exploitation" component

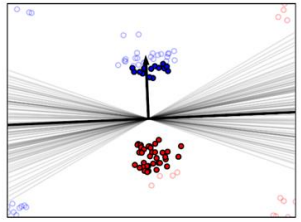
"Exploration" component



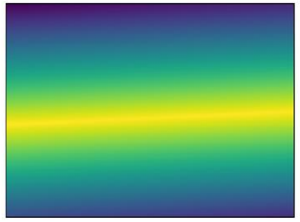
Labeled data pool



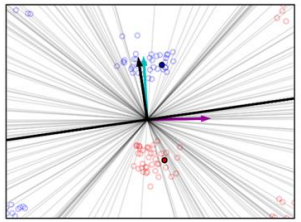
Uncertainty utility: initial



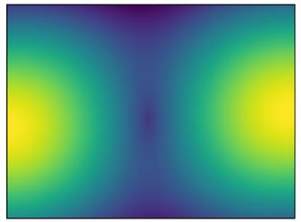
50 Uncertainty queries



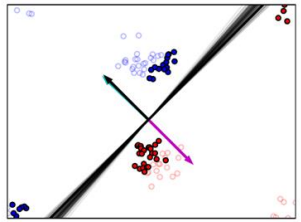
Uncertainty utility: 50 queries



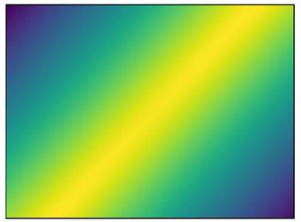
Hyperplane prior



APM-LR utility: initial



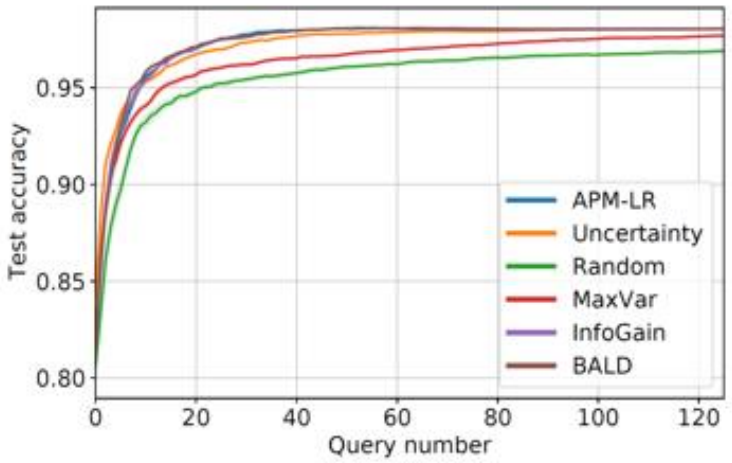
50 APM-LR queries



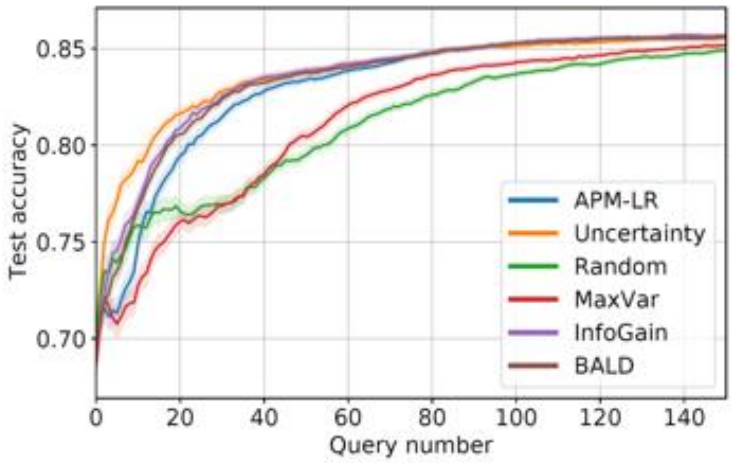
APM-LR utility: 50 queries

- Prevents sampling bias
- Theoretical guarantee on information gain

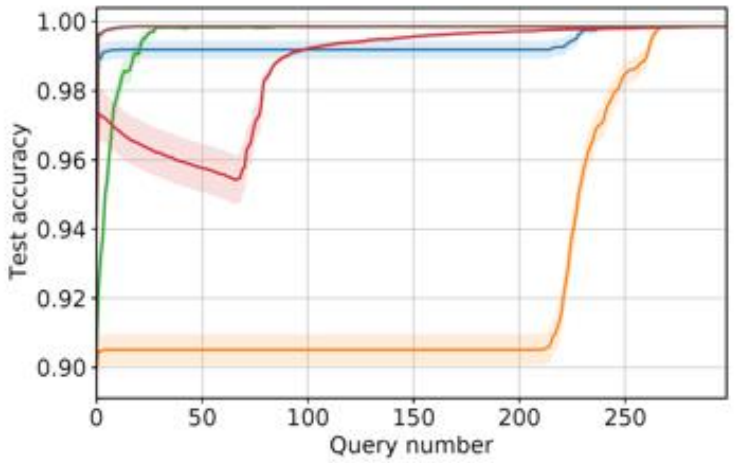
Experimental Results



letterDP



austra



cross

- Logistic regression active learning, UCI datasets (Dua and Graff, 2017)
- APM-LR comparable to other information maximization methods
- **Order of magnitude** cheaper computational cost than InfoGain, BALD (Houlsby et al. 2011)
- Outperforms uncertainty sampling in several cases (see paper for full results)

Conclusion

- APM converts information maximization problem to a geometric problem
- Balance between exploration and exploitation emerged naturally from first principles of feedback coding

Code available at: https://github.com/siplab-gt/APM-LR

http://siplab.gatech.edu

gregory.canal@gatech.edu

@GregHCanal