Optimal Teaching for Online Perceptrons

Xuezhou Zhang, Hrag Gorune Ohannessian, Ayon Sen, Scott Alfeld, Xiaojin Zhu

Department of Computer Sciences, University of Wisconsin-Madison

December 9, 2016
Optimal Teaching Problem

► **Student**: A machine learner $A$.
► **Teacher**: A person who knows a target model $\theta^*$, and wants to teach it to the student $A$ by creating a training set $D$.
► **Goal**: Find the ‘best’ training set.
► **General Optimization Formulation**:

\[
\min_D \text{loss}(A(D), \theta^*) + \text{effort}(D)
\]

Alternatively,

\[
\min_D \text{effort}(D), \quad \text{s.t.} \quad \text{loss}(A(D), \theta^*) \leq \epsilon
\]
Aim of this line of work

- Extend optimal teaching problem to **sequential** learners.
- Explore the teaching setting of **uncertainty**, i.e. the lack of information on the teacher’s side.
Motivating Example: Online Perceptrons

Algorithm 1 Online Perceptron

1: Learning parameters: Initial weight vector $w_0 \in \mathbb{R}^d$, learning rate $\eta > 0$.
2: \textbf{for} $t = 1 \ldots$ do
3: \hspace{1em} receive $x_t$
4: \hspace{1em} predict $\hat{y}_t = \text{sign}(\langle w_{t-1}, x_t \rangle)$
5: \hspace{1em} receive $y_t$
6: \hspace{1em} $w_t \leftarrow w_{t-1} + 1_{(y_t \langle x_t, w_{t-1} \rangle \leq 0)} \eta y_t x_t$

- **General Setting**: Allow non-zero $w_0$ and arbitrary learning rate $\eta$.
- **Formulation**: In this example, the machine learner $\mathcal{A}$ is the perceptron, and model $\theta$ is the linear decision boundary represented by the parameter $w$. 
Teaching with Full Knowledge of the Perceptron

Definition
The **Exact Teaching dimension** of perceptron is defined as

\[
\arg\min_{\mathcal{D}} |\mathcal{D}|,
\]

s.t. \( A(\mathcal{D}) = w^* \)

Theorem
*For any target parameter \( w^* \), a perceptron with any initial weight \( w_0 \) and learning rate \( \eta \) has exact teaching dimension 1.*
Approximate Teaching with Unknown W0

Definition
The $\epsilon$-Approximate Teaching dimension of perceptron is defined as

$$\arg\min_{D} |D|,$$

s.t. $$\frac{\langle A(D), w^* \rangle}{\|A(D)\|\|w^*\|} \geq 1 - \epsilon$$

Theorem
For any target parameter $w^*$ and precision $\epsilon$, a perceptron with unknown initial weight $w_0$ and known learning rate $\eta$ has $\epsilon$-approximate teaching dimension 3.
Discussion and Future Work

- An ‘interactive’ or ‘collaborative’ learning setting, where the student tries to learn the target model, while the teacher learns to teach.

- One potential solution is through active learning. Here the teacher can be formulated as an active learner who learns by probing the student and receive its feedback.

- However, this does not capture the interactive nature in teaching, and only optimizes the teacher’s learning task.

- Better solution?