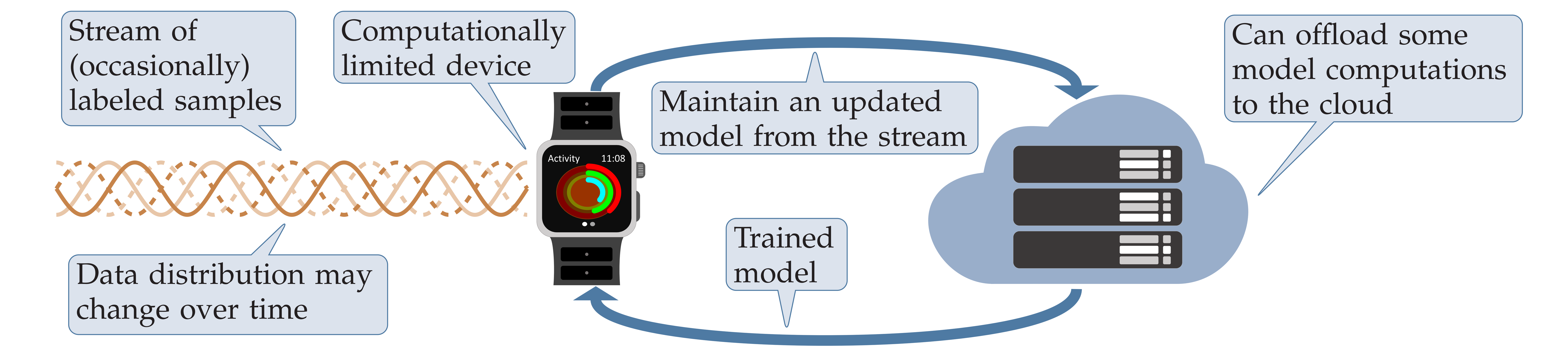
# Online Linear Models for Edge Computing

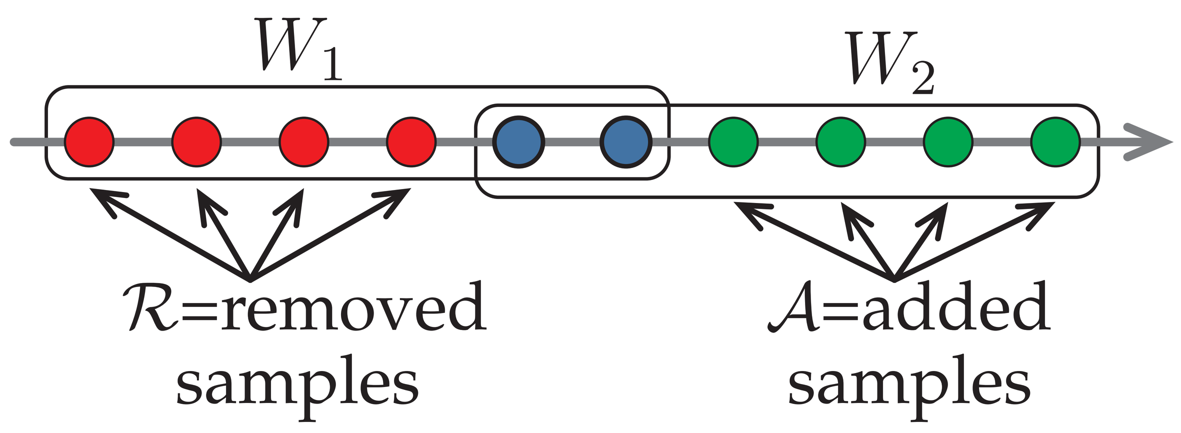

Hadar Sivan[1], Moshe Gabel[2], Assaf Schuster[1]

[1]Technion – Israel Institute of Technology, [2]University of Toronto


## The Setting



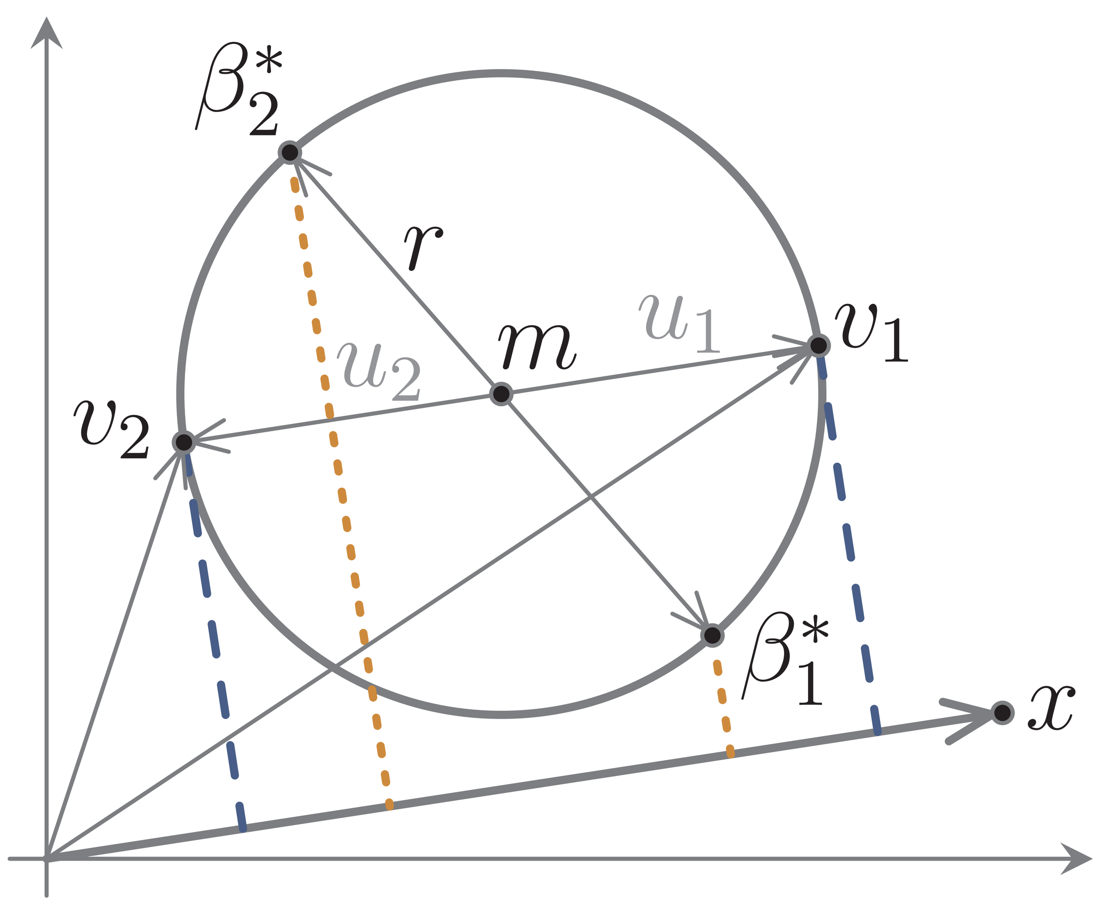

## Method and Tools




Bound the distance between the optimal models of the first window, $\beta_1^*$, and a second window, $\beta_2^*$, without computing $\beta_2^*$.

$\beta^* = \arg\min_{\beta} C \sum_{i \in W} \ell(y_i, x_i^T \beta) + \frac{1}{2}\|\beta\|^2$

is an $L_2$-regularized model with loss $\ell$.




**Theorem 1.** Let $\Delta g$ be

$$\Delta g := \boxed{\sum_{i \in \mathcal{A}} \nabla \ell_i(\beta_1^*)} - \boxed{\sum_{i \in \mathcal{R}} \nabla \ell_i(\beta_1^*)}.$$

Then:

$$\|\beta_1^* - \beta_2^*\| \le 2\|r\|,$$

where $r = \frac{1}{2}\left(\beta_1^* - \frac{C_2}{C_1}\beta_1^* + C_2 \Delta g\right)$.

$L_2$-Regularized LR: $r = \frac{C}{2}\Delta g$
Ridge Regression: $r = \frac{1}{4\alpha}\Delta g$

## Our Solution

### DRUiD – Drift detectoR from boUnded Distance

- Trades off a small number of batch model computations for better accuracy.
- Suitable for linear model with $L_2$-regularized convex differentiable loss.
- Supports both classification and regression tasks.

Monitors the **distance** between the **last batch-computed model** ($\beta_1^*$) and the **hypothetical model** that could be computed from the current position of the sliding window ($\beta_2^*$).

## DRUiD Algorithm

**For every new labeled sample:**

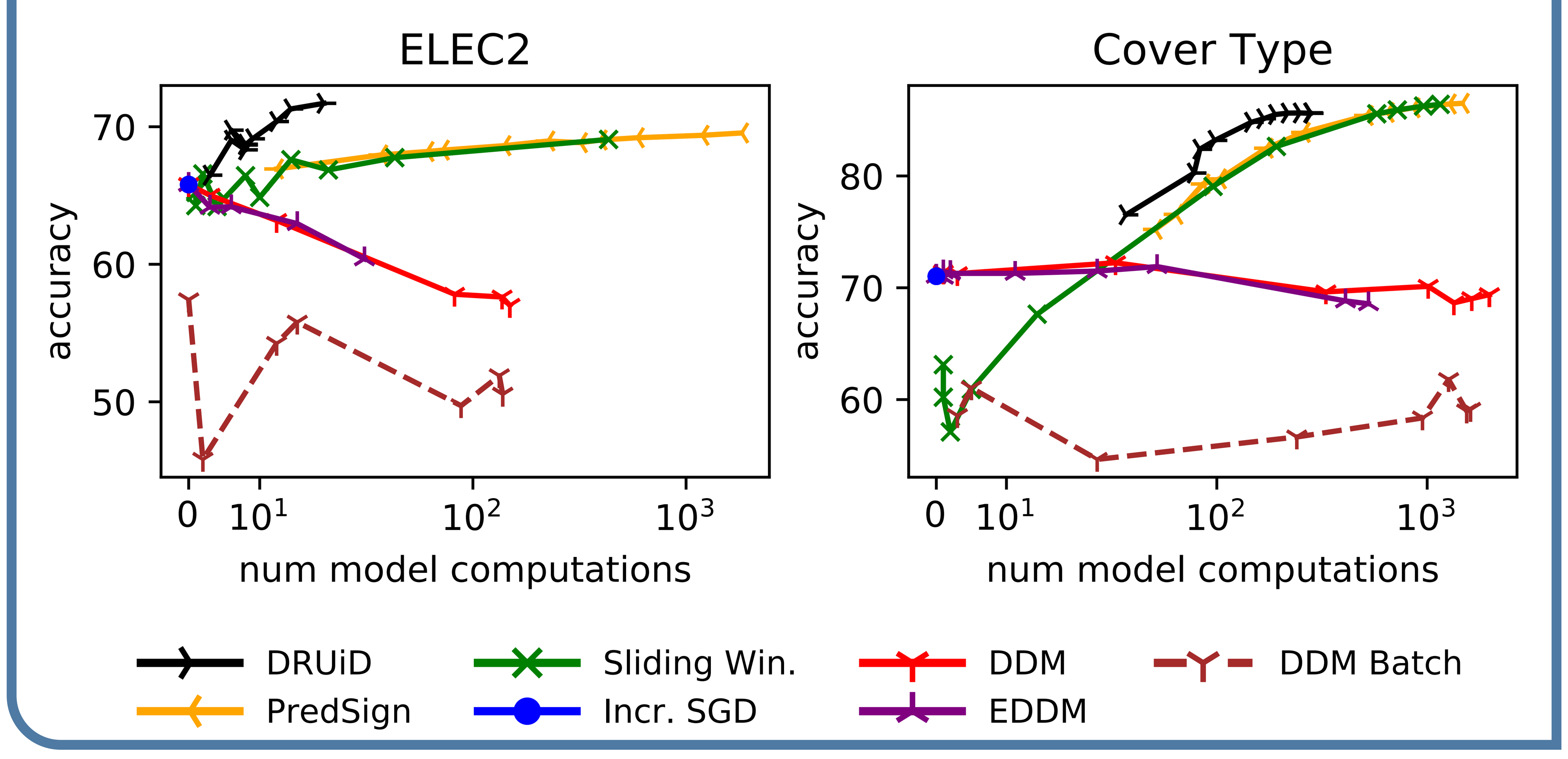
1. Update the sliding window $W$
2. Update $\|r\|$ using Theorem 1
3. If $\|r\| > \mathit{Threshold}$, compute $\beta_1^*$ from the samples in $W$

**Prediction for a new sample $x$:**

1. Return the prediction of $\beta_1^*$ for $x$

## Results on Real Data

DRUiD achieves better tradeoff of accuracy vs number of computations than existing algorithms.




## Results on Ill-Conditioned Problems

DRUiD's batch mode performs well on ill-conditioned problems, while incremental algorithms perform poorly.