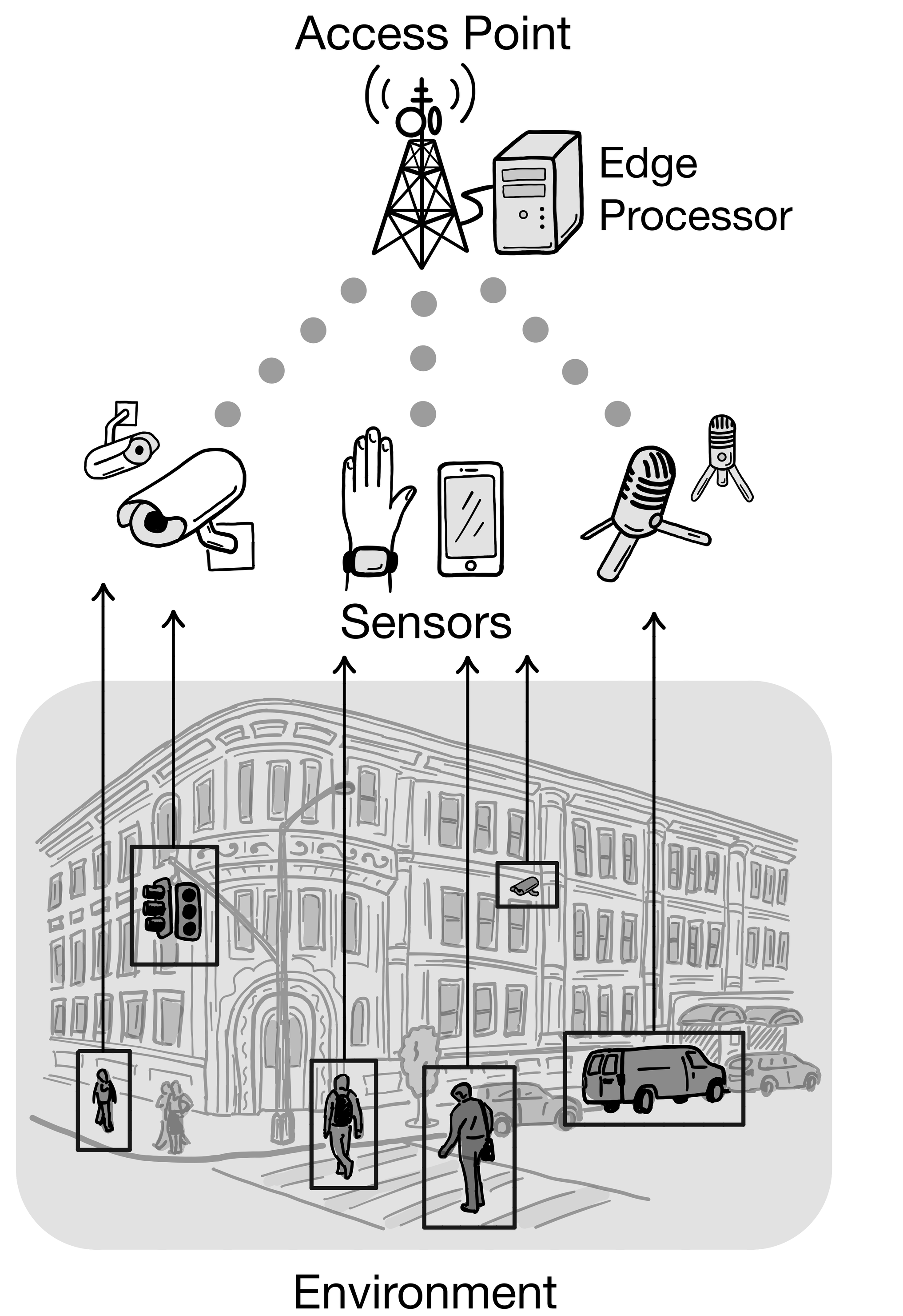
Edge-Assisted On-Sensor Information Selection for Bandwidth-Constrained Systems

Igor Burago   Marco Levorato

Department of Computer Science, University of California, Irvine

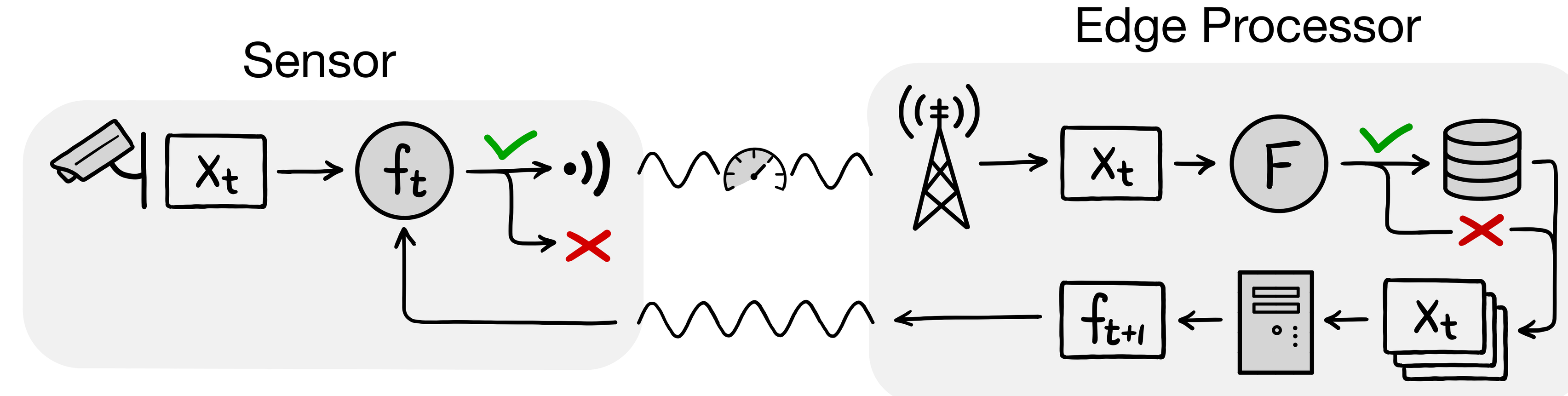


Introduction

Autonomous adaptation in IoT systems has unique requirements for implementing the induced learning and control tasks, atypical for a classic ML setting. Data access, model updates, and decision making processes often cannot be all implemented within a single (powerful) computing device.

Typical ML		IoT
Single computing unit	Data access → Processing → Decision making	Multiple communicating devices

The data sources — wireless **sensors** — are quite constrained due to their mobility both in computing power and available bandwidth resources. Hence, they can neither do the full processing locally, nor transmit all of their observations to (or beyond) the **edge** of the wireless island — the only places fully capable of the required processing necessary for the system to react to its changing environment.



Consider the pair of a single sensor and an edge processor, tasked with detecting observations of interest $x_t \in X^{(1)}$ in the sensor's data stream $\{x_t\}$. The rest $x_t \in X^{(0)}$ are to be ignored. The reference decision $F: x_t \in X^{(1)}$ can only be computed at the edge.

Problem 1: No Bandwidth Optimization

— A simplified decision rule $f(x_{t+1}, \theta_t) \gtrless \eta_t$ is maintained by the edge and offloaded to the sensor.
— The edge sees an uncensored sample of $\{x_t\}$ (e.g., there is no bandwidth constraints).

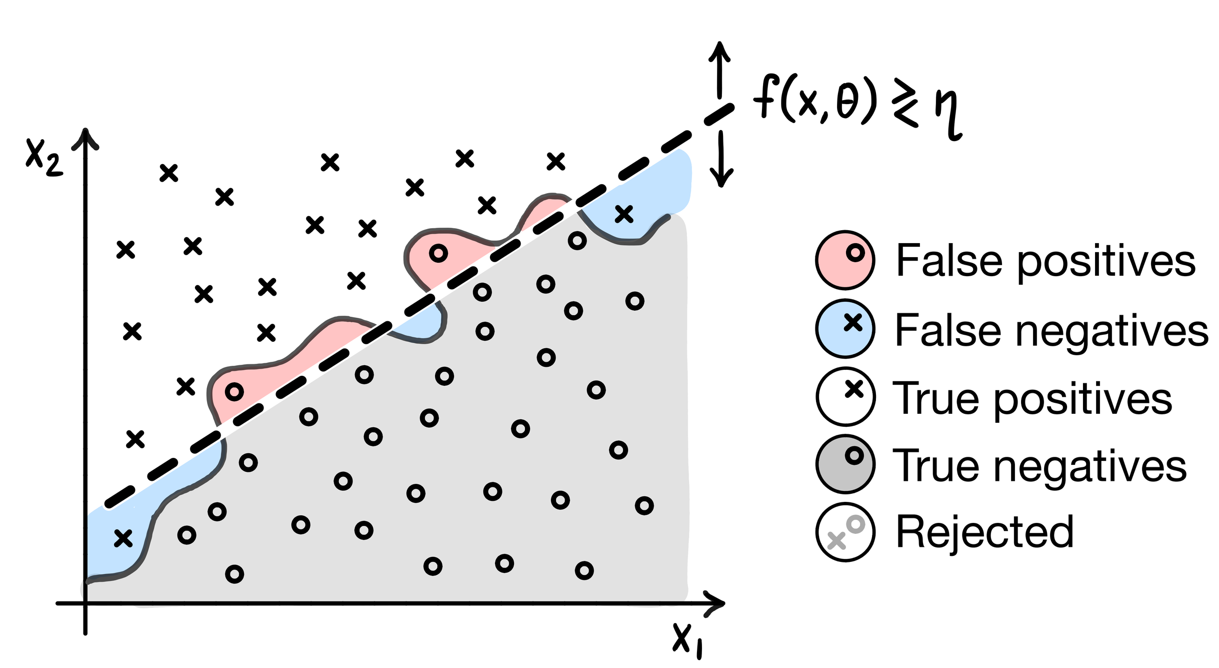
Criteria:

(1) Balanced error probabilities:
$\pi_0 \alpha(\theta,\eta) = \pi_1 \beta(\theta,\eta)$.

(2) Minimax error probability
$\alpha(\theta,\eta) = \beta(\theta,\eta)$.

(3) Neyman-Pearson:
$\beta(\theta,\eta) \longrightarrow \min_{\theta,\eta}$, s.t. $\alpha(\theta,\eta) \leq \alpha_*$.

Here the function f structurally defines the decision with parameters θ and threshold η; α and β are the false positive and false negative probabilities, respectively; $I^{(i)}(x) = \mathbb{1}[x \in X^{(i)}]$ are the supervision indicators, and $\hat{I}^{(i)}_{\theta,\eta}(x) = \mathbb{1}[f(x,\theta) > \eta]$ is the detection indicator; π_0, π_1 are a priori probabilities.



For fixed parameters θ_t, threshold η_t is updated at every time slot t, as follows:

(A) $\eta_{t+1} = \eta_t + \gamma_t \big(I^{(0)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1}) - I^{(1)}(x_{t+1}) \hat{I}^{(0)}_t(x_{t+1})\big)$;

(B) $\eta_{t+1} = \eta_t + \gamma_t \big(\hat{\pi}^{(1)}_{t+1} I^{(0)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1}) - \hat{\pi}^{(0)}_{t+1} I^{(1)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1})\big)$;

(C) $\eta_{t+1} = \eta_t + \gamma_t \big(I^{(0)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1}) - \alpha_0\big)$, $\alpha_0 = \pi_0 \alpha_*$.

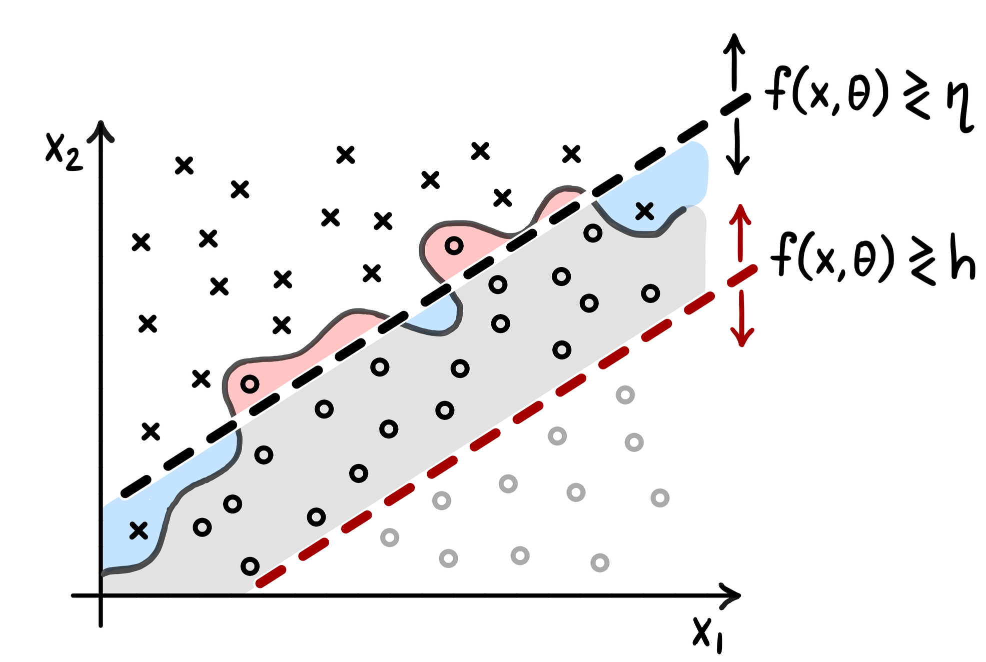
Parameters θ_t are updated at the same or less frequent schedule using the Kiefer-Wolfowitz procedure, each time with the growing sample of observations accumulated at the edge.

Theorem 1
For independent $\{x_t\}$;
$\gamma_t > 0$, $\sum \gamma_t = \infty$, $\sum \gamma_t < \infty$;
p.d.f. of $f(x,\theta)$ being smooth and nonzero outside of optimality region:
(a) Algorithm (A) achieves criterion (1) w.p. 1.
(b) Algorithm (B) achieves criterion (2) w.p. 1.
(c) Algorithm (C) achieves criterion (3) w.p. 1.

Problem 2: With Bandwidth Optimization

— A simplified decision rule $f(x_{t+1}, \theta_t) \gtrless \eta_t$ is maintained by the edge.
— The edge sees a sample of $\{x_t\}$ censored with a supplementary decision rule $f(x_{t+1}, \theta_t) \gtrless h_t$ which is offloaded to the sensor.

Criterion (4): Neyman-Pearson criterion (3) with an additional constraint $\alpha(\theta, h) = \varphi_* \ (> \alpha_*)$.



For fixed parameters θ_t, thresholds η_t, h_t are updated as follows:

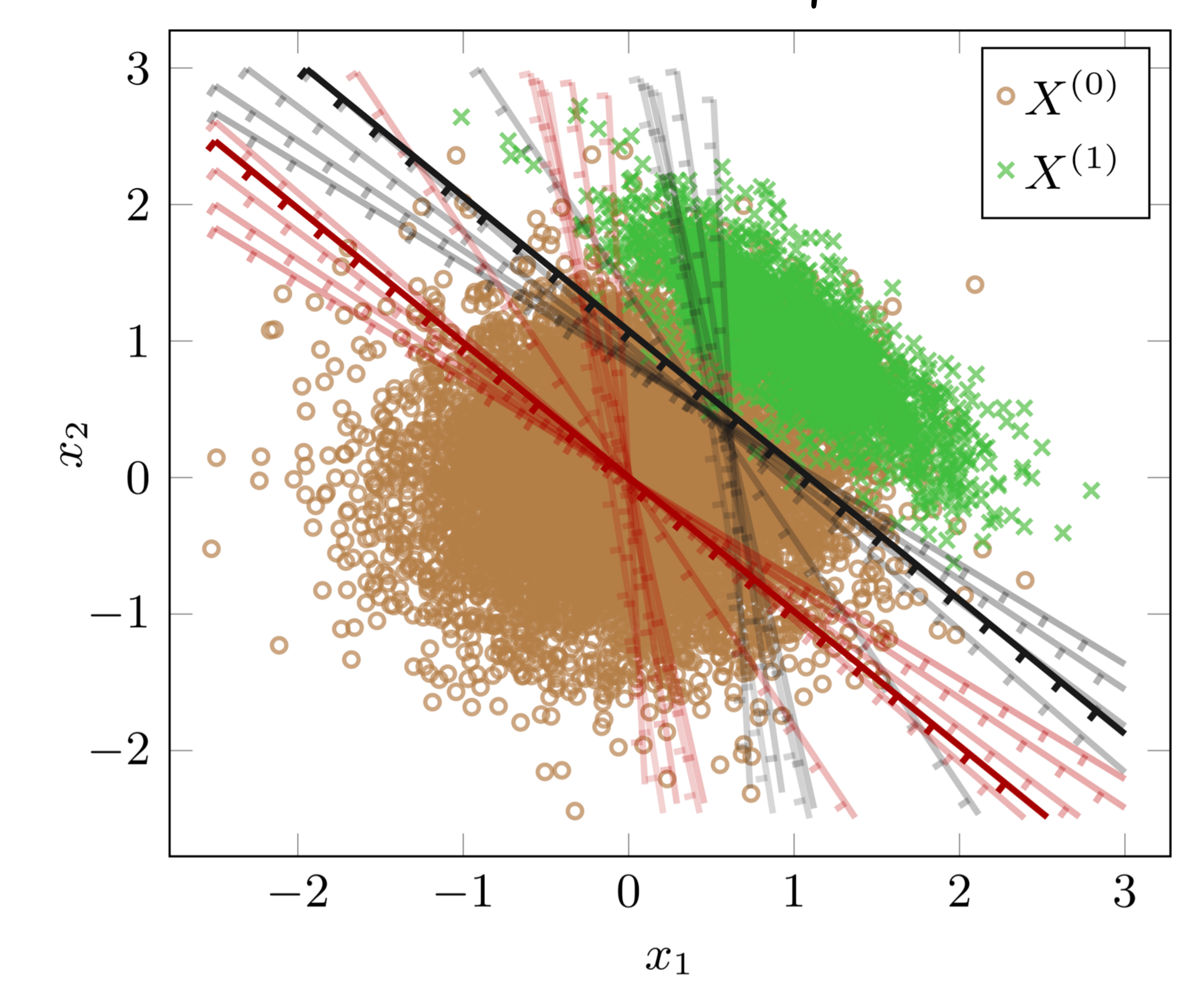
$\tilde{\eta}_{t+1} = \eta_t + \gamma_t \big(I^{(0)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1}, \eta_t) - \pi_0 \alpha_*\big)$,
(D) $\tilde{h}_{t+1} = h_t + \gamma_t \big(I^{(0)}(x_{t+1}) \hat{I}^{(1)}_t(x_{t+1}, h_t) - \pi_0 \varphi_*\big)$,
$(\eta_{t+1}, h_{t+1}) = P_D(\tilde{\eta}_{t+1}, \tilde{h}_{t+1})$,

where P_D is the projector onto $D = \{(\eta, h) : h \leq \eta\}$.

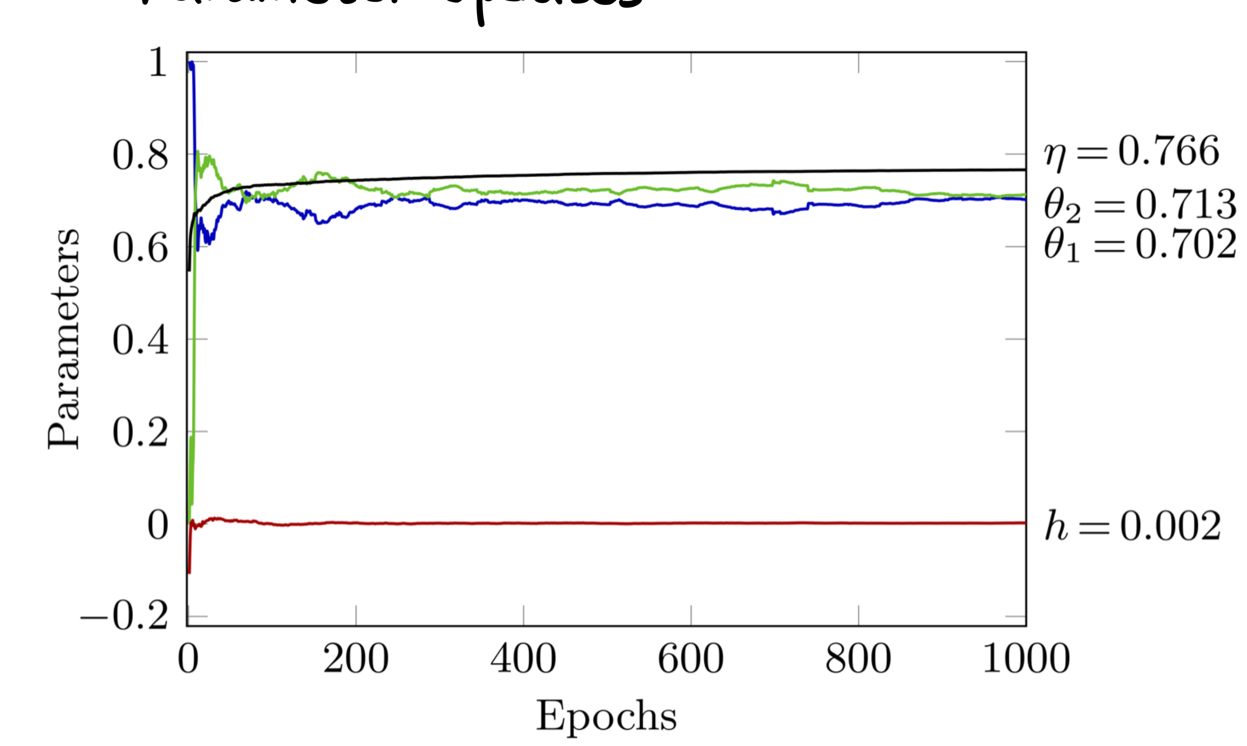
Theorem 2
Under the conditions of Theorem 1,
Algorithm (D) achieves criterion (4) w.p. 1.

Example

Separating surface updates
$f(x,\theta_t) = \eta_t$ and $f(x,\theta_t) = h_t$
for an exemplary two-dimentional data distributions with $\alpha_* = 0.1$, $\varphi_* = 0.5$:



Parameter updates



Every 100 slots constitute an epoch, during which vector θ_t remains unchanged.
After last epoch: $\alpha = 0.11$, $\beta = 0.02$, $\varphi = 0.50$.

2018-02-14 · San Diego, CA