Parameter-Free Convex Learning through Coin Betting
Francesco Orabona and Dávid Pál
Yahoo Research, NY
Are You Still Tuning/Learning/Adapting Hyperparameters?

Standard Machine Learning procedures

Regularized empirical risk minimization:

$$\arg\min_{w \in \mathbb{R}^d} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{N} f(w, x_i, y_i)$$

where $f$ is convex in $w$. 
Are You Still Tuning/Learning/Adapting Hyperparameters?

Standard Machine Learning procedures

Regularized empirical risk minimization:

\[
\arg \min_{w \in \mathbb{R}^d} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{N} f(w, x_i, y_i)
\]

where \(f\) is convex in \(w\).

- How do you choose the regularizer weight \(\lambda\)?
Are You Still Tuning/Learning/Adapting Hyperparameters?

Standard Machine Learning procedures

Stochastic approximation:

\[ w_t = w_{t-1} - \eta_t \nabla f(w_{t-1}, x_t, y_t) \]

where \( f \) is convex in \( w \).
Are You Still Tuning/Learning/Adapting Hyperparameters?

Standard Machine Learning procedures

Stochastic approximation:

$$w_t = w_{t-1} - \eta_t \nabla f(w_{t-1}, x_t, y_t)$$

where $f$ is convex in $w$.

- How do you choose the learning rate $\eta_t$?
Wasn’t machine learning about learning automatically from data?

- There is a history of 7 years of parameter-free algorithms that *do not have learning rates nor regularizers to tune*.
- But they were very unintuitive and complex
One Coin to Rule Them All

Online Coin betting algorithms give rise to *optimal* and *parameter-free* learning algorithms
Simple Algorithm & Good Results

- Parameter-free
- Extremely simple algorithm
- Same complexity of SGD
- Kernelizable

See how at the poster!