# Dimension-Free Exponentiated Gradient a.k.a. Gradient Descent without any Parameter

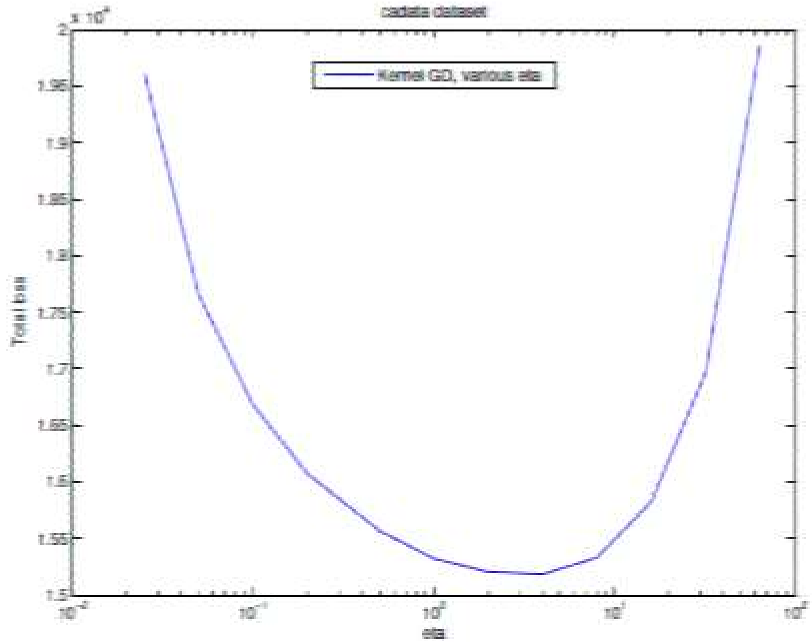
Francesco Orabona, Toyota Technological Institute at Chicago

What is Online Learning?

- Sequential prediction over rounds $t = 1, 2, \ldots$
- At each round t
  - Learner receives a question $\boldsymbol{x}_t \in \mathbb{X}$
  - Predicts answer $\hat{y}_t$
  - Receives "correct" answer $y_t \in Y$ and suffers loss $\ell(\hat{y}_t, y_t)$
- Allows $x_t$ and $y_t$ to be generated by an adversary
- Goal of learner: minimize "Regret" over $T$ rounds
- Online learning "implies" stochastic optimization

# Unconstrained Online Learning with Lipschitz Loss

- Learning rate should be $\frac{\eta}{\sqrt{t}}$, e.g. Zinkevich (2003)
- Setting $\eta$ optimally the regret is $\mathcal{O}(\|\boldsymbol{u}\|\sqrt{T})$ otherwise $\mathcal{O}((\|\boldsymbol{u}\|^2+1)\sqrt{T})$
- Unfortunately the optimal $\eta$ depends on the *future*



# An Improved Lower Bound

Bad news: The optimal rate $\mathcal{O}(\|\boldsymbol{u}\|\sqrt{T})$ *cannot* be achieved!

**Theorem**

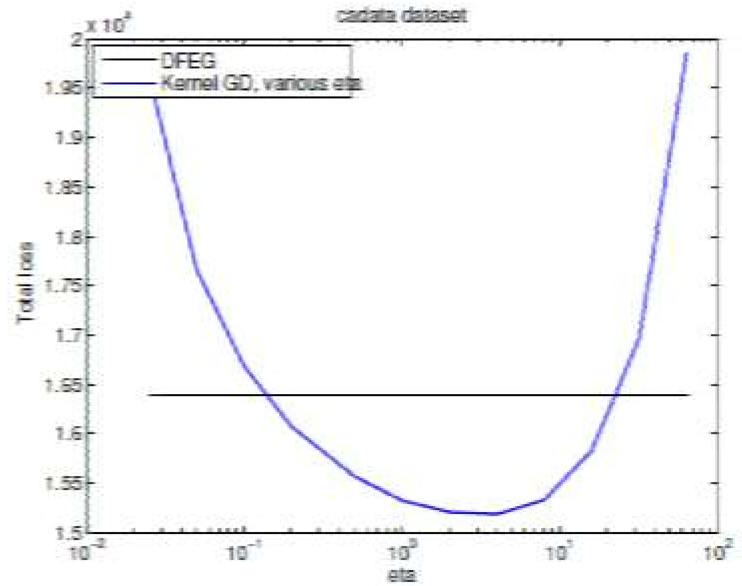
- *Regret against* $\boldsymbol{u} = 0$ *is 0* $\Rightarrow$ *Regret against other* $\boldsymbol{u}$ *is* $\Omega(T)$
- *Regret against* $\boldsymbol{u} = 0$ *is* $\epsilon > 0 \Rightarrow$ *Regret against other* $\boldsymbol{u}$ *is* $\Omega\left(\|\boldsymbol{u}\|\sqrt{T}\sqrt{\log\frac{\|\boldsymbol{u}\|\sqrt{T}}{\epsilon}}\right)$

Is it possible to obtain this regret without having to tune any parameter?

See also Streeter and McMahan (2012)

# Dimension-Free Exponentiated Gradient

- FTRL with a new regularizer, and new analysis
- Parameter-free
- Same complexity of SGD
- Dimension-Free ⇒ Kernelizable




**Theorem**

*For any sequence of Lipschitz convex losses*

$$Regret(T) \leq \mathcal{O}\left(\|\boldsymbol{u}\|\sqrt{T}\left(\ln\left(T^{1.5}\|\boldsymbol{u}\|\right)-1\right)\right)$$

See how at poster Sun13

Do you like adaptive algorithms? Read my other paper too: Kpotufe&Orabona, "Regression-tree Tuning in a Streaming Setting"