ANALYSIS OF HEDGE ALGO.

\[ \text{Potential: } P_t = -\ln \sum_i w_0,i \cdot e^{\eta L_{t,i}} \]

\[ \text{Due to normalization} \]

\[ P_t - P_{t-1} = -\ln \sum_i w_0,i \cdot e^{-\eta L_{t,i}} + \ln \sum_i w_{t,i} \cdot e^{-\eta L_{t-1,i}} \]

\[ = -\ln \frac{\sum_i w_0,i \cdot e^{-\eta L_{t-1,i}} \cdot e^{-\eta L_{t,i}}}{\sum_i w_0,i \cdot e^{-\eta L_{t-1,i}}} \]

\[ = -\ln \sum_i w_{t-1,i} \cdot e^{-\eta L_{t,i}} \]

\[ \geq -\ln \sum_i w_{t-1,i} \left( 1 - (1 - e^{-\eta}) L_{t,i} \right) \]

\[ e^{-\eta} \times \xrightarrow{x \in [0,1]} 1 - (1 - e^{-\eta}) \times \]

\[ = -\ln \left( \sum_i w_{t-1,i} - (1 - e^{-\eta}) \cdot \frac{x}{w_{t-1,i} \cdot L_{t,i}} \right) \]

\[ \geq (1 - e^{-\eta}) \cdot \bar{w}_{t-1} \cdot \bar{L}_t \]

DROP OF POTENTIAL

\[ \geq (1 - e^{-\eta}) \text{ LOSS OF ALGO.} \]
\[
\sum_{t=1}^{T} p_t - p_{t-1} \geq (1 - e^{-\eta}) \sum_{t=1}^{T} w_{t-1} L_t
\]

Lower bound

\[
\sum_{t=1}^{T} p_t - p_{t-1} = p_T - p_0 \to 0
\]

\[
\ln \sum_{i} w_{0,i} e^{-\eta} L_{T,i}
\]

\[
\leq \ln w_{0,i} e^{-\eta} L_{T,i}
\]

\[
= - \ln w_{0,i} + \eta L_{T,i}
\]

Upper bound

\[
\sum_{t=1}^{T} w_{t} L_{t} \leq \frac{\ln w_{0,i} + \eta L_{T,i}}{1 - e^{-\eta}}
\]

L_{ALG}

If \( \bar{w}_i = \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \) THEN \( \ln \frac{1}{\bar{w}_i} = \ln n \)

CAN HANDLE LOTS OF EXPERTS

\[
L_{ALG} \leq \frac{1}{1 - e^{-\eta}} \ln n + \frac{\eta}{1 - e^{-\eta}} L_{T,i}
\]

\( \eta = 1.58 \ln n + 1.58 L_{T,i} \)

\( \uparrow \) WANT 1

IF \( \eta \) TUNED AS FUNCTION OF \( n \) & \( \hat{L} \) THEN

\[
\sum_{t=1}^{T} w_{t} L_{t} \leq \frac{\ln n + \sqrt{2 \ln n + \ln n}}{\hat{L}}
\]

IF \( L^{*} \leq \hat{L} \) REGRET BOUND
BIG PICTURE

- We used Exponential Weights and Cohmin to achieve regret bounds.

- Expected loss bounds hold for arbitrary sequences.

- Expectation w.r.t. internal randomization of Alg.

- Logarithmic dependence on # of experts, typical for "multiplicative" updates.

QUESTIONS:

- Lower bounds?

- Motivation of updates?

- Where did the potential come from?

- What about other loss functions?

- Compare against best linear combination of experts?
Lots of "stupid" experts are "specialized" combined to something better.

Later: Boosting
- Iteratively builds small linear combination of weak hypotheses.