ANALYSIS OF HEDGE ALG.

\[ \text{POTENTIAL: } P_t = -\ln \sum_i w_{0,i} e^{-\eta L_{t,i}} \]

\[ \text{DUE TO NORMALIZATION} \]

\[ P_t - P_{t-1} = -\ln \sum_i w_{0,i} e^{-\eta L_{t,i}} + \ln \sum_i w_{t,i} e^{-\eta L_{t-1,i}} \]

\[ = -\ln \frac{\sum_i w_{0,i} e^{-\eta L_{t,i}} e^{-\eta L_{t,i}}}{\sum_i w_{0,i} e^{-\eta L_{t-1,i}}} \]

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

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

\[ e^{-\eta x} \leq 1 - (1-e^{-\eta}) x \]

\[ x \in [0,1] \]

\[ = -\ln \left( \sum_i w_{t-1,i} - (1-e^{-\eta}) \bar{w}_{t-1} \cdot \bar{L}_t \right) \]

\[ \ln(1-x) \leq -x \]

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

DROP OF POTENTIAL

\[ \geq (1-e^{-\eta}) \text{ LOSS OF ALG.} \]
\[ \sum_{t=1}^{T} P_t - P_{t-1} \geq (1 - e^{-\eta}) \sum_{t=1}^{T} w_{t-1} \cdot L_t \]
\[ \sum_{t=1}^{T} P_t - P_{t-1} = \frac{P_T - P_0}{w_0} \]
\[ = - \ln \frac{\sum_{i=1}^{L} w_{0,i} e^{-\eta L \leq t, i}}{1} \]
\[ \leq - \ln w_{0,i} e^{-\eta L \leq t, i} \]
\[ = - \ln w_{0,i} + \eta L \leq t, i \]

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

If \( \bar{w}_i = \left( \frac{1}{n} \ldots \frac{1}{n} \right) \) THEN \( \ln \frac{1}{w_{0,i}} = \ln n \)
- CAN HANDLE LOTS OF EXPERTS

\[ L_{\text{alg}} \leq \frac{1}{1-e^{-\eta}} \ln n + \frac{\eta}{1-e^{-\eta}} L \leq t, i \] (\(*\))

\[ \eta = 1.58 \ln n + 1.58 L \leq t, i \]
\[ \uparrow \text{WANT 1} \]

IF \( \hat{L} \geq L^* \) AND \( \eta = \sqrt{\frac{2 \ln n}{n}} \) THEN

\[ \text{REGRET BOUND} \]

\[ (*) \leq \min \frac{L \leq t, i}{{L^*}} + \sqrt{2} \ln n + \ln n \]
BIG PICTURE

- We used exponential weights and softmax 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?
SO FAR

MASTER

WEIGHTED MAJORITY

EXPONENTIAL WEIGHTS

$E_1, E_2, E_3, E_n$

LOTS OF "STUPID" EXPERTS ARE "SPECIALIZED"
COMBINED TO SOMETHING BETTER

LATER: BOOSTING
- ITERATIVELY BUILDS
  SMALL LINEAR COMBINATION
  OF WEAK HYPOTHESIS