NOISE FREE CASE: 3 CACHE THAT GETS ALL HITS & NO MISSES

WEIGHTED MEDIAN ALC.
(DETERMINISTIC):

\[ M_{\text{ALC}} \leq \log_2 n \]
\[ \beta = 0 \]

NOISY CASE: \( M^* = \min_i M_i \)
\[ \text{MISSES OF } i\text{-TH CACHE} \]
\[ M_{\text{ALC}} \leq \frac{\log n + \ln \frac{1}{\beta} M^*}{1 + \beta} \]
\[ \leq 2 M^* + 2 \sqrt{M^* \ln n} + \log_2 n \]
\[ \beta \]
\[ \text{TUNED} \]
\[ \# \text{ OF CACHES IS } n \]
PROBABILISTIC ALG.
- USE EXPONENTIAL WEIGHT $w_i \sim \beta^{M_i}$
- CHOOSE i W. PROB. Wi

(RANDOMISE WEIGHTED MAJORITY ALG)

$$E(M_{\text{ALL}}) \leq \frac{\ln n + (\ln \beta) M_i}{1 - \beta}$$

$$\leq M^* + \sqrt{2M^* \ln n} + \ln n$$

TUNED $\beta$

NOISE-FREE CASE:

$$E(M_{\text{ALL}}) \leq \ln n$$

TOO MUCH REFETCHING

- IN PARTICULAR THE PROBABILISTIC ALG.
LOWER BOUND:

THAT ANY DET. ALG CAN BE FORCED TO MAKE $10\log_2 n$ MISTAKES IN NOISE-FREE CASE

PROOF SKETCH:

\[
\text{RANGE OF CACHES WITH NO MISES} \quad \Rightarrow \quad \text{CURRENT HYPOTHESIS}
\]

ONE OF THESE TWO FILES CAUSE THE RANGE TO BE CUT DOWN BY $\leq \frac{1}{2}$

\[
n^{\frac{1}{2}} \leq 1
\]

\[
10\log_2 n - M_{\text{ALG}} \leq 0
\]

\[
M_{\text{ALG}} \geq 10\log_2 n
\]

LOWER BOUND FOR NOISY CASE

\[
M_{\text{ALG}} \geq 2M^* + \mathbb{E} \left( \frac{\sqrt{M^* \ln n}}{\ln n} \right) + \ln n
\]

\[
\mathbb{E}(M_{\text{ALG}}) \geq M^* + \ldots
\]
\[ M((c_1, c_2, \ldots, c_k)) = \sum_{i=1}^{k} M_i c_i \quad \text{TOTAL # OF MISSES IN LIST } i \]

\[ w(\bar{c}) \sim \beta^i \quad \text{TOTAL # OF MISSES OF ANY CACHE WITH } \bar{c} \text{ IN LIST } i \]

\[ \sim \prod_{i} \beta^{M_i c_{i+1}} \quad \text{PER TRIAL EACH CACHE 0 OR 1 MISS} \]

\[ \sim n^k \text{ CACHES} \]

HANDLING EXPONENTIALLY MANY CACHE W. DYNAMIC PROBR.

\[ W(i, s) := \text{ALL PARTIAL CACHES} \]

\[ \sum_{0 \leq k, 0 \leq j \leq n} (c_1, \ldots, c_i), \text{ s.t.} \]

\[ \sum_{j=1}^{i} c_j = s \]

\[ W(i, s) = \sum_{q=0}^{s} W(i-1, s-q) \beta^{M_i c_{i+1}} \]

\[ O(n^2 k) \quad \text{EXPENSIVE?} \]
IDEA FOR DETERMINISTIC ALG?

\[ V(i, c) = \text{TOTAL WEIGHT OF ALL CACHES GOING THRU GAP C IN LIST}\]

\[ \text{COMPUTE WEIGHTED MEDIAN OF } V(i,*) \]

USE THE K MEDIAN AS CACHE

BOUNDS SHOULD MULTIPLY BY K

PROBLEM: RESULTING CACHE CAN BE TOO BIG

\[
\begin{array}{ccccccc}
  m & = & 5 \\
  k & = & 4 \\
  \beta & = & \frac{1}{2} \\
  & & 3 \cdot 3 & 3 \cdot 3 & 2 \cdot 6 & 0 \cdot 8 \\
  & & 0 & 0 & 4 \cdot 4 & 4 \cdot 8 = M(4,2) \\
  & & 0 & 0 & 0 = M(2,3) & 0 & 0 & 4 \cdot 4 \\
  \text{MEDIAN CACHE SIZE} & + & 2 & & & & 0
\end{array}
\]
LOWER BOUND FOR DET. ALG. SHOULD BE
\geq \text{const.} \cdot k \cdot 19^2 \frac{m}{n} \quad \text{for noise-free case}

ADVERSARY STRATEGY: CHOOSE A MISS THAT PRODUCES MINIMUM REDUCTION OF RANGE

PROBABILISTIC ALG:

USE RANDOMIZED WEIGHTED MAJORITY ALG ON ALL \approx m^k CACHES

- PICK RANDOM CACHE

\[ E(M_{\text{alg}}) \leq M^* + \sqrt{2 M^* k \ln n} + k \ln n \]

BEST OF \approx m^k

- TOO EXPENSIVE \( O(n^2 k) \)

WANT \( O(n k) \) ALG

- TOO MUCH REFETCHING
Sketch of proofs for shifting experts

\[ w_{t+1}^m = w_{t+1} e^{-m L_{t+1}} \]

\[ w_{t+1} = \alpha \left( \frac{1}{n} \right) + (1 - \alpha) w_{t}^m \]

Fixed share to start vector

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

Sum of unnorm weights

Potential: \[ P_t = -\ln W_t \]

We showed

\[ w_t \cdot L_t \left( 1 - e^{-\eta} \right) \leq P_t^m - P_t \]

\[ = -\ln \frac{w_t^m}{W_t} \]

Also: \[ w_{t+1} \leq w_t^m \]

\[ 1 \leq \frac{w_t^m}{W_{t+1}} \]

\[ 0 \leq \ln \frac{W_t}{W_{t+1}} = -\ln \frac{w_{t+1}^m}{W_t} \]

Sum

\[ \sum_{t=1}^{T} w_t L_t \leq \frac{\ln \frac{w_{T+1}^m}{W_t}}{1 - e^{-\eta}} \]
TOTAL WEIGHT OF ALL PARTITIONS

\[ W_{t+1} \geq \text{ANY FIXED PARTITION } P \text{ WITH } K \text{ SHIFTS} \]

\[ \frac{1}{n} e^{-\eta \text{ LOSS}(P)} \frac{1}{(1-\alpha)^{t-k-1}} \left( \frac{\alpha}{m} \right)^k \]

\[ \sum_{t=1}^{T} \tilde{W}_t \tilde{C}_t \leq \frac{1}{1-e^\eta} \left( \ln n + \eta \text{ LOSS}(P) + k \left( \ln \frac{T-k}{n} + \ln n + (k-1-k) \ln \left(1 + \frac{k}{T-k-1} \right) \right) \right) \]

\[ \approx \frac{1}{1-e^\eta} \left( \# \text{ OF BITS TO ENCODE PARTITION} + \eta \text{ LOSS}(P) \right) \]

FANCIER VARIABLE SHARE ANALYSIS

IDEA BEHIND ALG:
- EXPERTS THAT DO WELL DON'T SHARE WEIGHTS
- IF AN EXPERT HAS HIGH LOSS
  THEN SOME OF ITS WEIGHT \( \rightarrow \) POOL
  POOL IS SHARE
Loss Update

\[ W_{t,i} \rightarrow W_{t+1,i} \]

\[ W_{t,i} \hat{m} (1-\alpha) \overset{\text{LTI}}{\rightarrow} \text{POOL} \]

\[ W_{t,i} \hat{m} (1-(1-\alpha)^{L_t,i}) \]

\[ W_{t+1,i} = W_{t,i} \hat{m} (1-(1-\alpha)^{L_t,i}) + \text{SHARE OF POOL} \]

Lots of Loss
Shares to \( E_t \)

\[ E_5 \]

Small Loss
\( E_t \) gets enough weight to recover

Lots of Details

See original papers
LONG TERM MEMORY $u$ SHIFTS WITHIN $u$ 10

$\text{ (FIXED SHARE TO UNIF. PAST) }$

$$\text{LOSS UPDATE } \bar{W}_t$$

$$\bar{W}_{t+1} = \alpha \bar{v}_t + (1-\alpha) \bar{W}_t$$

$$\bar{v}_t = \frac{\sum_{q=1}^{t-1} W_q^m}{t-1}$$

- WITHIN SEGMENT $(1-\alpha)$
- NEW SEGMENT $\alpha/m_t$
- CONTINUATION SEGMENT $\alpha/m_t$

MORE TRICKY

$$L_{LAC} \leq \frac{1}{1-c_n} \left( \eta \text{ LOSS (P)} + \# \text{ OF BITS TO ENCODE P} \right)$$

$\ln n$ BITS FOR NEW EXPERT

$\ln n$ " " CONTINUATION
NEW SIGNATURE PROBLEM
CO-LEARNING

K TASKS EACH SOLVED BY 1 OF N EXPERTS

1 8 15
2 9
3 10
4 11
5 12
6 13
7 14

TASKS ARE PARTITIONED INTO M GROUPS
BEST EXPERT SAME IN EACH OF THE GROUPS

N >> m

TRIVIAL BOUND

TOTAL LOSS ≤ \frac{\kappa \ln N + \eta \sum L_i}{1 - e^{-\eta}}

WANT \kappa \ln N REPLACED BY

\kappa \ln M + m \ln \frac{N}{m}

LONG TERM MEMORY METHODOLOGY?
IT CAN DO IT IF ROW BY ROW