Today

Multiplicative Weight Update Method (part IV)

- Approximately solving LPs
- Boosting in learning
- Hardcore Set Lemma

CS5.205.1

Toolkit in TCS

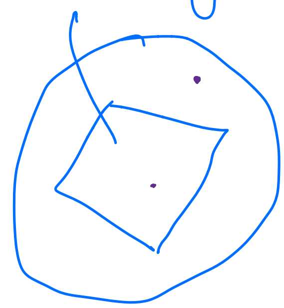
- Lecture #18

(21 Apr '21)

Instructor: Prahladh Harsha

Approximately Solving LPs

$Ax \geq b$ $\mathcal{P}$



$\exists ?\ \underbrace{x \in \mathcal{P}}_{\text{easy}},\ \underbrace{Ax \geq b}_{\text{hard.}}$

Reduce to single constraint case where the single constraint is a convex combination of the m constraints in "$Ax \geq b$"

$Ax \geq b$

$$\left.\begin{array}{l} A_1 x \geq b_1 \ \text{---}\ p_1 \\ A_2 x \geq b_2 \ \text{---}\ p_2 \\ \vdots \\ A_m x \geq b_m \ \text{---}\ p_m \\ p_i \geq 0 \quad \sum p_i = 1 \end{array}\right\} \Rightarrow p^T A x \geq p^T b.$$

(ℓ,ρ) - Bounded Oracle

$0 \le \ell \le \rho$. There exists an oracle $\mathcal{O}$ & a set $I \subseteq [m]$ that on input p - prob dist either

(1) Declares $p^T A x \ge p^T b$ is infeasible

or

(2) Finds a feasible soln x s.t

$$\forall x \in I, \quad -\ell \le A_i x - b_i \le \rho.$$

$$\forall x \notin I, \quad -\rho \le A_i x - b_i \le \ell.$$

ρ - width of oracle.

Approximate LP solver

1. Initialize $p^{(1)} = (\frac{1}{m}, \dots \frac{1}{m})$.
2. For $t \leftarrow 1..$ to T
   - Use oracle to find a feasible soln to
     $$p^{(t)^T} \cdot Ax \ge p^{(t)^T} \cdot b.$$
   - If oracle declares 'infeasible' output "infeasible" & exit

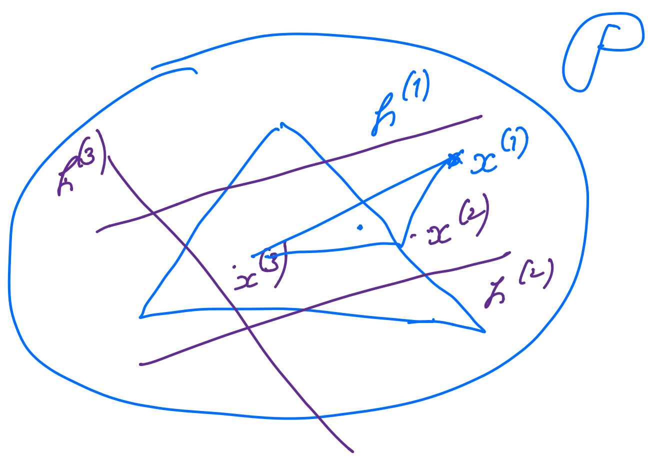
▷ Else get feasible soln $x^{(t)}$

▷. Construct Loss vector $M^{(t)}$ as follows

$$m_i^{(t)} \leftarrow \frac{[A_i x^{(t)} - b_i]}{\rho}$$

▷. Update $p^{(t+1)} \leftarrow p^{(t)}$ & $m^{(t)}$ & learning parameter ε

3. Output $\tilde{x} = \frac{1}{T}\sum_{t=1}^{T} x^{(t)}$



Thm [Approx LP Solver] $\forall\, 0 \le \delta \le 1$

Approx LP-solver finds a δ-additive approximate feasible solve $\tilde{x}$

i.e, $\forall i \in [m]$ $\quad A_i \tilde{x} \ge b_i - \delta$

or.

outputs "$Ax \ge b$" is infeasible

using at most $O\left(\frac{l\rho \log m}{\varepsilon^2}\right)$ calls to oracle & $O(m)$ processing time per oracle call

Pf: $m_i^{(t)} = \frac{(A_i x^{(t)} - b_i)}{\rho}$ | Assume alg does not declare infeasibility

Cost of the MWUM at round t:

$= \langle M^{(t)}, p^{(t)} \rangle = \frac{1}{\rho} \sum_{i=1}^{m} \left(p^{(t)^T} A_i x^{(t)} - p^{(t)^T} b_i \right)$

≥ 0

$A_i x - b_i \geq 0$

For $i \in I$

$m_i^{(t)} = \frac{(A_i x^{(t)} - b_i)}{\rho}$

MWUM thm.

$$\sum_{t=1}^{T} \langle M^{(t)}, p^{(t)} \rangle \leq \sum_{t=1}^{T} m_i^{(t)} + \varepsilon \sum_{t=1}^{T} |m_i^{(t)}| + \frac{\ln m}{\varepsilon}$$

$$0 \leq \sum_{t=1}^{T} \frac{(A_i x^{(t)} - b_i)}{\rho} + \varepsilon \sum_{t=1}^{T} \frac{|A_i x^{(t)} - b_i|}{\rho} + \ln m / \varepsilon$$

$$\leq (1+\varepsilon) \sum_{t=1}^{T} \frac{(A_i x^{(t)} - b_i)}{\rho} +$$

$$2\varepsilon \sum_{t:\ <0} \frac{|A_i x^{(t)} - b_i|}{\rho}$$

$$+ \frac{\ln m}{\varepsilon}$$

$$\leq (1+\varepsilon) \sum_{t=1}^{T} \frac{(A_i x^{(t)} - b_i)}{\rho} + \frac{2\varepsilon l T}{\rho} + \frac{\ln m}{\varepsilon}$$

$$0 \leq (1+\varepsilon) \frac{1}{T} \sum_{t=1}^{T} (A_i x^{(t)} - b_i) + 2\varepsilon l + \frac{\rho \ln m}{\varepsilon T}$$

$$\tilde{x} = \frac{1}{T} \sum x_i^{(t)}$$

$$0 \leq (1+\varepsilon)(A_i \tilde{x} - b_i) + 2\varepsilon l + \frac{\rho \ln m}{\varepsilon T}$$

$$\varepsilon = \frac{\delta}{4l} \quad ; \quad T = \frac{8 l \rho \ln m}{\delta^2}$$

$$0 \leq (1+\varepsilon)(A_i \tilde{x} - b_i) + \delta$$

$$A_i \tilde{x} \geq b_i - \frac{\delta}{1+\varepsilon} \geq b_i - \delta.$$

For $i \notin I$, analysis is similar. $\boxtimes$

Remarks

1. Constraint Matrix.

$$Ax \geq b.$$

- Entries of A are non-negative } - covering. $I = [m]$
- Entries of A are non-positive. $A'x \leq b'$ (A' - non-negative entries) } - packing $I = \phi$

2. Width - not nice

eg. weighted set cover [Garg - Konneman]

3. Approximate Oracle is sufficient

4. Linearity of constraints ??

Linearity used to conclude

$$A_i \tilde{x} - b_i \geq \frac{1}{T} \sum A_i x^{(t)} - b_i.$$

But any convex constraints satisfies the above.

(ie, the method applies to convex feasible solns).

Application III: Boosting in Learning Theory

PAC Learning: Probably Approximately Correct Learning

$\mathcal{X}$ - finite domain $\quad \{0,1\}^n$

$c: \mathcal{X} \to \{0,1\}$ (concept)

Access to c is via random queries

$$\left.\begin{array}{l} (x_1, c(x_1)) \\ (x_2, c(x_2)) \\ \vdots \\ (x_N, c(x_N)) \end{array}\right\} \Longrightarrow \boxed{\begin{array}{c}\text{Learner} \\ L\end{array}} \Longrightarrow h: \mathcal{X} \to \{0,1\}$$

$x_i \sim \mathcal{D}$ on $\mathcal{X}$

$\mathcal{D}$ - distribution on $\mathcal{X}$

Error of hypothesis

$$\mathbb{E}_{x \sim \mathcal{D}}\left[\,|h(x) - c(x)|\,\right]$$

Strong Learning: Concept C is strongly learnable if $\exists$ a learner L

$\forall$ dist $\mathcal{D}$ on X and all $\mu, \delta \in (0,1)$

$$\Pr_L [\, err(h) > \mu \,] \le \delta$$

- Learner L to run in time poly $\frac{1}{\mu}, \frac{1}{\delta}, \log|X|$

γ-Weakly learning

$\forall$ dist $\mathcal{D}$ $\quad \forall \delta$.

$$\Pr_L [\, err(h) > \tfrac{1}{2} - \gamma \,] \le \delta.$$

Qn:

Given a γ-weak learner $\rightarrow$ can we obtain a strong learner?

Boosting: YES.

Thm (MWUM - in terms of relative entropy)

Assuming $m_i^{(t)} \in [0,1]$ & $\varepsilon \in (0, 1/2]$.

Expected cost

$$\sum_{t=1}^{T} \langle M^{(t)}, p^{(t)} \rangle \le (1+\varepsilon) \sum_{t=1}^{T} \langle M^{(t)}, P \rangle + \frac{RE(P \| p^{(1)})}{\varepsilon}$$

for any fixed dist P on experts & $p^{(1)}$ - starting distribution

Key Insight: Weak learner works against all dist $\mathcal{D}$ (not just the dist $\mathcal{D}$ we care for).

$$\left.\underbrace{\begin{matrix} (x_1, c(x_1)) \\ (x_2, c(x_2)) \\ \vdots \\ (x_N, c(x_N)) \end{matrix}}_{\text{training set } S}\right\} \Rightarrow \boxed{L'} \Rightarrow h$$

Fix the training set S of N samples.

Typically Weak learner puts the uniform dist on S.

Across rounds of MWUM

Change dist on training set s.t
dist puts more wt on incorrectly classified inputs.

MMUM:

Experts: N elts from the training set
$x_1 \ldots x_N$

$p^{(t)}$ – on N samples

$p^{(1)}$ – uniform dist.

$p^{(t)} \rightarrow$ Weak learner $\rightarrow h^{(t)}$

Cost $\quad m_x^{(t)} = 1 - |h^{(t)}(x) - c(x)|$

Run MWUM for T rounds
& ε - learning parameter.

$h^{(1)}, \ldots \quad h^{(T)}$

$\tilde{h} = \text{majority}(h^{(1)} \ldots h^{(T)})$

Expected Loss (across T rounds) of MWUM alg

$$\langle M^{(t)}, p^{(t)} \rangle \geq \frac{1}{2} + \gamma$$ (Guarantee of Weak learner)

$$\sum_{t=1}^{T} \langle M^{(t)}, p^{(t)} \rangle \geq \left(\frac{1}{2} + \gamma\right) T.$$

E = incorrectly classified i/ps in S

$$= \{ x \in S \mid \tilde{h}(x) \neq c(x) \}$$

$\forall x \in E$

$$\sum_{t=1}^{T} m_x^{(t)} \leq T/2$$

$$\sum_{t=1}^{T} \langle M^{(t)}, p^{(t)} \rangle \leq (1+\varepsilon) \sum_{t=1}^{T} \langle M^{(t)}, p \rangle + \frac{RE(p \| p^{(1)})}{\varepsilon}$$

Use $p^{(1)}$ = uniform dist

p - uniform dist on E.

$$\left(\frac{1}{2} + \gamma\right) T \leq (1+\varepsilon) \frac{T}{2} + \frac{\ln(N/|E|)}{\varepsilon}.$$

$$RE(P \| Q) = \sum p_i \ln \frac{p_i}{q_i}$$

$$\varepsilon := \gamma \quad ; \quad T = \left\lceil \frac{2}{\gamma^2} \ln\left(\frac{1}{\mu}\right) \right\rceil$$

$$\frac{\gamma T}{2} \le \frac{\ln(N/|E|)}{\gamma}$$

$$\Rightarrow \frac{|E|}{|N|} \le \mu \qquad \text{since } T \ge \frac{2}{\gamma^2} \ln\left(\frac{1}{\mu}\right)$$

$\tilde{h}$ classifies at most μ fraction of inputs in S incorrectly

$\boxtimes$

Next lecture:

- Boosting to give a constructive of hardcore set lemma.
- MWUM to Matrix-valued costs.