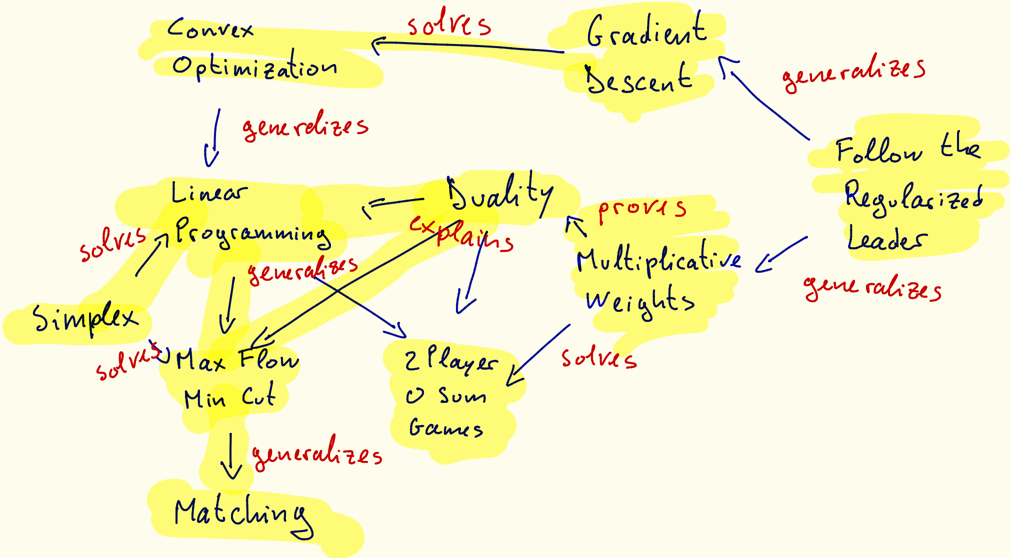
Follow The Regularized Leader


Convex Optimization
solves
Gradient Descent
generalizes
generelizes
Follow the Regularized Leader
Linear Programming
Duality
proves
explains
solves
generalizes
Multiplicative Weights
Simplex
generalizes
solves
Max Flow Min Cut
2 Player 0 Sum Games
solves
generalizes
Matching

Time	Algorithm	Loss	Loss of algorithm
$t=1$	$x_1^{(1)}, \dots, x_n^{(1)}$ $x_1^{(1)} + x_2^{(1)} + \dots + x_n^{(1)} = 1$ $x_i^{(1)} \geq 0$	$\ell_1^{(1)} \dots \ell_n^{(1)}$	$\sum_i x_i^{(1)} \ell_i^{(1)}$
$t=2$			

Time	Algorithm	Loss	Loss of al
$t=1$	$x^{(1)} \in K$	$f_1 : K \to \mathbb{R}$	$f_1(x^{(1)})$
$t=2$	$x^{(2)} \in K$	$f_2 : K \to \mathbb{R}$	$f_2(x^{(2)})$
	$\vdots$		

Follow the Leader Algorithm

at time t

$$x^{(t)} = \underset{x \in K}{\operatorname{argmin}} \; f_1(x) + \dots f_{t-1}(x)$$

$K =$ distributions $\qquad f_t(x) = \sum_i \ell_i^{(t)} x_i$

time	Alg	Loss
1	$(\frac{1}{2}, \frac{1}{2})$	$(0, \frac{1}{2})$ $[f_1(x) = \frac{1}{2} x_2]$
2	$(1, 0)$	$(1, 0)$
3	$(0, 1)$	$(0, 1)$
4	$(1, 0)$	$(1, 0)$

Analysis of FTL

Theorem

$$\text{Regret}_T = \sum_{t=1}^{T} f_t(x^{(t)}) - \min_{x \in K} \sum_{t=1}^{T} f_t(x)$$

$$\leq \sum_{t=1}^{T} f_t(x^{(t)}) - f_t(x^{(t+1)})$$

Proof Induction on T

we want to prove

$$\sum_{t=1}^{T} f_t(x^{(t+1)}) \leq \min_{x \in K} \sum_{t=1}^{T} f_t(x) \quad (*)$$

$T = 1$ $\qquad f_1(x^{(2)}) \leq \min_{x \in K} f_1(x)$

✓ definition of $x^{(2)}$

$$x^{(2)} = \arg\min_{x \in K} f_1(x)$$

✓ Assume $(*)$ for T

$$\sum_{t=1}^{T+1} f_t(x^{(t+1)}) = f_{T+1}(x^{(T+2)}) + \sum_{t=1}^{T} f_t(x^{(t+1)})$$

$$\leq f_{T+1}(x^{(T+2)}) + \sum_{t=1}^{T} f_t(x^{T+2}) = \sum_{t=1}^{T+1} f_t(x^{T+2})$$

$$= \min_{x \in K} \sum_{t=1}^{T+1} f_t(x)$$

Follow the Regularized Leader Algorithm

Define a function $R: K \to \mathbb{R}$

Algorithm

$$x^{(t)} = \underset{x \in K}{\operatorname{argmin}}\ f_1(x) + \dots + f_{t-1}(x) + R(x)$$

Analysis of FTRL

Theorem: for every x

$$\left(\sum_{t=1}^{T} f_t(x^{(t)}) \right) - \left(\sum_{t=1}^{T} f_t(x) \right)$$

$$\leq \left(\sum_{t=1}^{T} f_t(x^{(t)}) - f_t(x^{(t+1)}) \right) + R(x) - R(x^1)$$

Proof

Imagine a game where FTL plays for $T+1$ rounds, with $R(\cdot)$ first cost function

$f_1 \dots f_T$ subsequent function

K = probability distribution

$f_t(x) = \sum_i \ell_i^{(t)} x_i$ for some $\ell_1^{(t)} \dots \ell_n^{(t)}$

Choose $R(\cdot)$

$\boxed{0 \log \frac{1}{0} = 0}$

Entropy of a distribution

$$H(x_1 \dots x_n) := \sum_{i=1}^{n} x_i \log \frac{1}{x_i}$$

$$H\left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) = 3 \cdot \left(\tfrac{1}{3} \log 3\right) = \log 3$$

$$H(0,1,0) = 1 \cdot \log 1 = 0$$

$$H\left(\tfrac{1}{3}, \tfrac{2}{3}, 0\right) = \tfrac{1}{3} \log 3 + \tfrac{2}{3} \log 1.5$$

$$0 \leq H(x) \leq \log n$$

$(0-1 \dots 0)$ $\qquad$ $\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$

$$R(x) = +c \cdot \sum_{i=1}^{n} x_i \ln x_i$$

Follow the Regularized Leader with Negative Entropy Regularizer

$$x^{(t)} = \underset{x \text{ prob. distrib.}}{\operatorname{argmin}} \; f_1(x) + \dots + f_{t-1}(x) + c_t \sum_{i=1}^{n} x_i \ln x_i$$

$$\min_{\substack{x:\ \sum_i x_i = 1 \\ x_i \geq 0}} \; \ell^{(1)} \cdot x + \ell^{(2)} \cdot x \dots + \ell^{(t-1)} \cdot x + c \sum_{i=1}^{n} x_i \ln x_i$$

$$f_\lambda = \ell^{(1)} \cdot x + \dots + \ell^{(t-1)} \cdot x + c \sum_{i=1}^{n} x_i \ln x_i + \lambda \cdot \left(\sum_i x_i - 1 \right)$$

$$\nabla f_\lambda(x) = \ell^{(1)} + \dots + \ell^{(t-1)} + c \cdot (\dots, \ln x_i + 1, \dots) + \lambda \cdot (1 \dots 1) = 0$$

$$x_i = e^{-\frac{1}{c}(\ell_i^1 + \dots \ell_i^{t-1})} \cdot e^{-\frac{\lambda}{c} - 1}$$

$$x_i = \frac{e^{-\frac{1}{c}(\ell_i^1 + \dots + \ell_i^{t-1})}}{\sum_{j=1}^{n} e^{-\frac{1}{c}(\ell_j^1 + \dots \ell_j^{t-1})}}$$

$$x_i^{(t)} = \frac{e^{-\frac{1}{c}(\ell_i^1 + \dots + \ell_i^{(t-1)})}}{\sum_{j=1}^{n} e^{-\frac{1}{c}(\ell_j^1 + \dots \ell_j^{(t-1)})}}$$

$$w_i^{(t)} = e^{-\frac{1}{c}(\ell_i^{(1)} + \dots \ell_i^{(t-1)})}$$

$$w_i^{(1)} = 1 \qquad w_i^{(t+1)} = \boxed{e^{-\frac{1}{c}\cdot \ell_i^{t}}} \cdot w_i^{(t)}$$

$$(1-\varepsilon)^{\ell_i^{(t)}}$$

$$x_i^{(t)} = \frac{w_i^{(t)}}{\sum_j w_j^{(t)}}$$

$$e^{-\frac{1}{c}} = (1-\varepsilon)$$

FTRL regret

$$\leq \sum_{t=1}^{T} \ell^{(t)} \cdot x^{(t)} - \ell^{(t)} \cdot x^{(t+1)}$$

$$+ R(x^{opt}) - R(x^{(1)})$$

$$\leq \sum_{t=1}^{T} \ell^{(t)} \cdot \left(x^{(t)} - x^{(t+1)}\right) + c \ln n$$

$$\leq \sum_{t=1}^{T} \ell^{(t)} \cdot \left(\frac{1}{c} \cdot x^{(t)}\right) + c \ln n$$

$$\leq T \cdot \frac{1}{c} \cdot 1 + c \ln n$$

$$\frac{T}{c} = c \ln n$$

$$c^2 = \frac{T}{\ln n}$$

$$= 2\sqrt{T \cdot \ln n}$$

$$c = \sqrt{\frac{T}{\ln n}}$$

Gradient Descent

$$K = \mathbb{R}^n$$

$f_t(x) = \ell_t \cdot x$ for some ℓ_t

$$R(x) = c \cdot \|x\|^2$$

$$x^{(t+1)} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \ \ell^{(1)} \cdot x + \cdots + \ell^{(t)} \cdot x + c \cdot \|x\|^2$$

$$\nabla \left(\sum_{i=1}^{t} \ell^{(i)} \cdot x + c\|x\|^2 \right)$$

$$= \ell^{(1)} + \cdots + \ell^{(t)} + 2cx$$

$$x^{(t+1)} = -\frac{1}{2c} \cdot \left(\ell^{(1)} + \cdots + \ell^{(t)} \right)$$

$$x^{(t+1)} = x^{(t)} - \frac{1}{2c} \cdot \ell^{(t)}$$

Regret

Suppose $g : \mathbb{R}^n \to \mathbb{R}$ convex, we want to minimize

When algorithm $x^{(t)}$ give loss function $\nabla g(x^{(t)})$