Об эффективности экспоненциального взвешивания в задачах о многоруких бандитах и выпуклой оптимизации

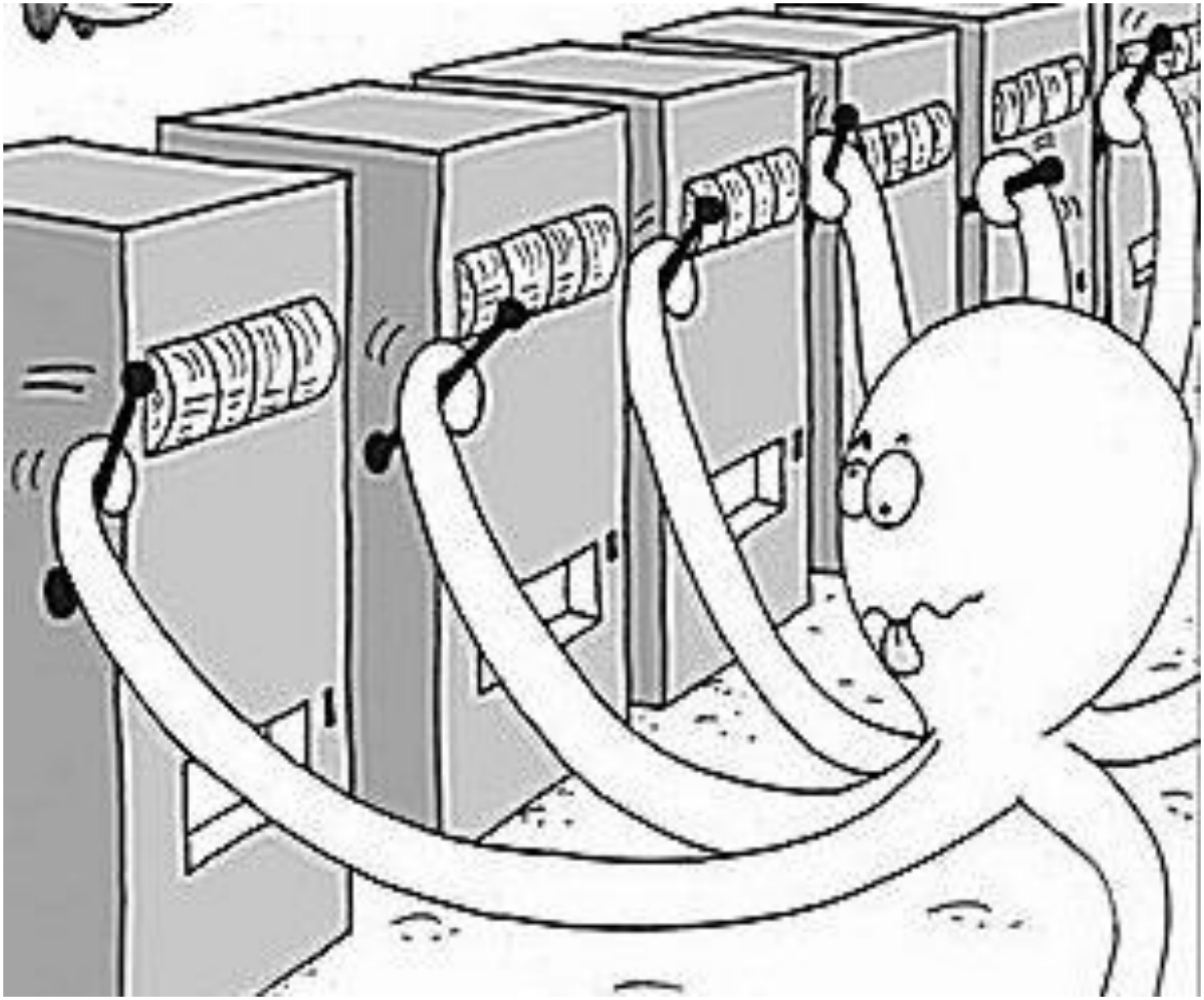
Александр Гасников
ПреМоЛаб, ФУПМ, МФТИ
avgasnikov@gmail.com

на основе совместных исследований с
Ю. Нестеровым (CORE UCL), В. Спокойным (WIAS),
В. Вьюгиным (ИППИ РАН), А. Назиным (ИПУ РАН)

Традиционная математическая школа V
Солнечногорск (Мос. обл.), 17 июня 2013

Многорукие бандиты

Имеется $n \gg 1$ различных ручек. “Дергание” каждой ручки приносит нам некоторые, вообще говоря, случайные потери r (*regret*). **Цель**: таким образом организовать процедуру дергания ручек (задается распределением вероятностей), чтобы ожидаемые суммарные потери после $N \gg 1$ дерганий (это число может быть заранее неизвестно!) были бы минимальны. **Проблема**: никакой априорной информации о ручках у нас нет ☹ Все, чем мы располагаем, это собственным опытом, полученным при дергании различных ручек (этого опыта нет на первом шаге). **Предположения**: математическое ожидание потерь не зависит от случайности в предыстории, но может меняться от шага к шагу, “потери” по абсолютной величине ограничены числом $\tilde{M}$.



Взвешивание экспертных решений

Имеется $n \gg 1$ различных Экспертов. Каждый эксперт играет на рынке, и все время проигрывает ☺ Пусть l_i^k – проигрыш Эксперта i на шаге k $(0 \le l_i^k \le M)$. На каждом шаге мы распределяем доллар между Экспертами. Потери, которые мы при этом несем, рассчитываются по потерям экспертов. **Цель**: таким образом организовать процедуру распределения доллара на каждом шаге, чтобы наши ожидаемые суммарные потери после $N \gg 1$ шагов (это число может быть заранее неизвестно!) были бы минимальны.

Отличие: от многоруких бандитов, в частности, в том, что здесь мы знаем всю историю игры, до настоящего момента.

Замечание 1: если потери эксперта i на шаге k, использующего стратегию ζ_i^k есть $\lambda\left(\omega^k, \zeta_i^k\right)$ $(0 \le \lambda(\cdot) \le M)$, и $\lambda(\cdot)$ – выпуклая функция по второму аргументу, то вместо распределения доллара между экспертами, мы выбираем стратегию – выпуклую комбинацию стратегий Экспертов.

Замечание 2: если при этом мы не можем гарантировать, что $\lambda(\cdot)$ – выпуклая функция по второму аргументу, то мы выбираем стратегию – распределение вероятностей на множестве стратегий Экспертов, и разыгрываем с.в. согласно этому распределению вероятностей. Причем ω^k может зависеть от нашего выбора распределения, но не может зависеть от того, что “выпадет” при разыгрывании!

Рекуррентное агрегирование оценок

Пусть имеется простая выборка $\xi = \xi^1, \xi^2, \ldots$ и функция потерь $E_\xi\left[f\left(x;\xi\right)\right]$, зависящая от вектора неизвестных параметров x из единичного симплекса. Огромный пласт задач машинного обучения сводится к поиску такого x, который доставляет минимум введенной функции. Однако есть серьезная проблема: распределение ξ, как правило, не известно, все что можно делать, это наблюдать независимые реализации ξ. Поэтому на практике (основываясь на з.б.ч.) минимизируют эмпирическую функцию потерь:

$$\frac{1}{N}\sum_{k=1}^{N} f\left(x;\xi^k\right).$$

Цель: имея реализации ξ, найти “оптимальный” x.

Рассмотрим задачу **стохастической оптимизации**

$$\frac{1}{N}\sum_{k=1}^{N} E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right] \to \min_{x\in S_n(1)},\ S_n(1)=\left\{x\geq 0:\ \ \sum_{i=1}^{n} x_i = 1\right\},\ \ (\heartsuit)$$

где ξ_k – независимые случайные величины (это условие можно существенно ослабить, см. далее условие (♪)),

$$E_{\xi^k}\left[\nabla_x f_k\left(x;\xi^k\right)\right] = \nabla_x E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right],$$

$f_k\left(x;\xi^k\right)$ – выпуклые функции по x, $\left\|\nabla_x f_k\left(x;\xi^k\right)\right\|_\infty \leq M$ – ограниченный градиент, это условие можно ослабить:

а) $E_{\xi^k}\left[\left\|\nabla_x f_k\left(x;\xi^k\right)\right\|_\infty^2\right] \leq M^2$; б) $E_{\xi^k}\left[\exp\left(\dfrac{\left\|\nabla_x f_k\left(x;\xi^k\right)\right\|_\infty^2}{M^2}\right)\right] \leq \exp(1)$.

Для решения задачи (♥) воспользуемся адаптивным методом зеркального спуска (в нашем случае: оптимизации на симплексе, методом экспоненциального взвешивания):

$$\frac{1}{N}\sum_{k=1}^{N} E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right] \to \min_{x\in S_n(1)}$$

Положим $x_i^0 = 1/n,\ i=1,\ldots,n$. Пусть $t=1,\ldots,N$.

Алгоритм МЗС-адаптивный

$$x_i^t = \frac{\exp\left(-\frac{1}{\beta_0\sqrt{t+1}}\sum_{k=1}^{t}\frac{\partial f_k\left(x^{k-1};\xi^k\right)}{\partial x_i}\right)}{\sum_{l=1}^{n}\exp\left(-\frac{1}{\beta_0\sqrt{t+1}}\sum_{k=1}^{t}\frac{\partial f_k\left(x^{k-1};\xi^k\right)}{\partial x_l}\right)},\quad i=1,\ldots,n,\quad \beta_0=\frac{M}{\sqrt{\ln n}}.$$

$$\underbrace{\left(E_{\xi,x}\left[f\left(\frac{1}{N}\sum_{k=1}^{N} x^{k-1};\xi \right)\right] - \min_{x\in S_n(1)} E_{\xi}\left[f\left(x;\xi\right)\right] \le \right)}$$

если все $f_k \equiv f$, а все ξ^k - независимы и одинаково распределены, как ξ

$$\frac{1}{N}\sum_{k=1}^{N} E_{\xi,x}\left[f_k\left(x^{k-1};\xi^k\right)\right] - \min_{x\in S_n(1)} \frac{1}{N}\sum_{k=1}^{N} E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right] \le 2M\sqrt{\frac{\ln n}{N}},$$

Это неравенство верно и при более слабых условиях:

а) $E_{\xi}\left[\left\| \nabla_x f\left(x;\xi\right)\right\|_{\infty}^{2} \right] \le M^2$ и

$$E_{\xi^k}\left\{ \nabla_x f_k\left(x^{k-1};\xi^k\right) - \nabla_x E_{\xi^k}\left[f_k\left(x^{k-1};\xi^k\right)\right] \middle| \xi^1,\ldots,\xi^{k-1} \right\} \overset{п.н.}{=} 0. \quad (♪)$$

$$P_{x^1,\ldots,x^N}\left\{\frac{1}{N}\sum_{k=1}^{N}E_{\xi^k}\left[f_k\left(x^{k-1};\xi^k\right)\right]-\min_{x\in S_n(1)}\frac{1}{N}\sum_{k=1}^{N}E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right]\geq\right.$$

$$\left.\geq\frac{2M}{\sqrt{N}}\left(\sqrt{\ln n}+\sqrt{3\Omega}\right)\right\}\leq\exp\left(-\Omega\right),$$

$$P_{x^1,\ldots,x^N}\left\{E_{\xi}\left[f\left(\frac{1}{N}\sum_{k=1}^{N}x^{k-1};\xi\right)\right]-\min_{x\in S_n(1)}E_{\xi}\left[f\left(x;\xi\right)\right]\geq\right.$$

$$\left.\geq\frac{2M}{\sqrt{N}}\left(\sqrt{\ln n}+\sqrt{3\Omega}\right)\right\}\leq\exp\left(-\Omega\right).$$

Это неравенство верно и при более слабом условии (♪); если вместо ограниченности градиента требовать только выполнение условия б), то $\sqrt{\ln n}+\sqrt{3\Omega}\to 12\sqrt{\ln n}+3\Omega.$

Мотивация 1

Введем *прокси-функцию* и *расстояние Кульбака–Лейблера*:

$$V(x) = \ln n + \sum_{i=1}^{n} x_i \ln x_i,\ V(x,y) = \sum_{i=1}^{n} x_i \ln\left(x_i / y_i\right) \geq \frac{1}{2}\|x-y\|_1^2,$$

последнее неравенство называют *неравенством Пинскера*.

Введем функцию (играющую роль функции Ляпунова для исследуемого итерационного процесса):

$$W_\beta(y) = \beta \ln\left(\frac{1}{n}\sum_{i=1}^{n} \exp\left(-\frac{y_i}{\beta}\right)\right).$$

Несложно проверить, что $W_\beta(y)$ – β-сопряженная для $V(x)$.

Приведем способ получения детерминированного МЗС.

$$x^t = \arg\min_{x \in S_n(1)} \left\{ \sum_{k=1}^{t} f_k\left(x^{k-1}\right) + \left\langle \sum_{k=1}^{t} \nabla f_k\left(x^{k-1}\right), x - x^{t-1} \right\rangle + \beta_t V\left(x\right) \right\},$$

$$x_i^t = -W_{\beta_t}\left(\sum_{k=1}^{t} \nabla f_k\left(x^{k-1}\right) \right),$$

или, что то же самое

$$x^t = \arg\min_{x \in S_n(1)} \left\{ f_t\left(x^{t-1}\right) + \left\langle \nabla f_t\left(x^{t-1}\right), x - x^{t-1} \right\rangle + \beta_t V\left(x, x^{t-1}\right) \right\}.$$

Аналогично можно показать, что стохастический вариант метода зеркального спуска может быть также представлен:

$$\begin{cases} y^k = y^{k-1} + \gamma_k \nabla_x f_k\left(x^{k-1}; \xi^k\right) \\ x^k = -\nabla W_{\beta_k}\left(y^k\right) \end{cases}, \ \gamma_k \equiv 1, \ \beta_k = \frac{M}{\sqrt{\ln n}} \sqrt{k+1}.$$

Мотивация 2

Аппроксимируя

$$\min_{x\in S_n(1)}\sum_{k=1}^{t} f_k\left(x\right)\approx \min_{x\in S_n(1)}\sum_{k=1}^{t}\left\{ f_k\left(x^{k-1}\right)+\left\langle \sum_{k=1}^{t}\nabla f_k\left(x^{k-1}\right), x-x^{t-1}\right\rangle\right\},$$

полагая при этом

$$P\left(x_j^t=1;\, x_i^t=0, i\neq j\right)\overset{def}{=}P_\varepsilon\left(\, j=\arg\max_{i=1,\ldots,n}\sum_{k=1}^{t}\left\{\left[-\nabla f_k\left(x^{k-1}\right)\right]_i+\varepsilon_{k,i}\right\}\right),$$

получим, при довольно общих условиях относительно с.в. $\varepsilon_{k,i}$ (типа i.i.d.), что если $t \gg 1$, то

$$P_{\varepsilon}\left(j = \arg\max_{i=1,\ldots,n} \sum_{k=1}^{t} \left\{ \left[-\nabla f_k\left(x^{k-1} \right) \right]_i + \varepsilon_{k,i} \right\} \right) \approx$$

$$\approx P_{\varsigma}\left(j = \arg\max_{i=1,\ldots,n} \left\{ \sum_{k=1}^{t} \left[-\nabla f_k\left(x^{k-1} \right) \right]_i \right\} + \varsigma_{t,i} \right),$$

с.в. $\varsigma_{t,i}$ – i.i.d. с распределением Гумбеля (max-устойчивым), с параметром, зависящим от t: $P\left(\varsigma_{t,i} < \tau \right) = \exp\left\{ -e^{-\tau/\beta_t} \right\}$. Тогда

$$E_{\varsigma}\left[x^t \right] = -W_{\beta_t}\left(\sum_{k=1}^{t} \nabla f_k\left(x^{k-1} \right) \right).$$

Ключевая выкладка

- Немировский–Юдин, 1979 !!!
- Поляк–Юдицкий, 1992; Немировский и др., 1999
- Нестеров; Юдицкий–Назин–Цыбаков–Ваятис, 2005
- Немировский–Юдицкий–Шапиро–Лан, 2009–2012

$$\sum_{k=1}^{t}\gamma_k\left\{E_{\xi,x}\left[f_k\left(x^{k-1};\xi^k\right)\right]-E_{\xi,x}\left[f_k\left(x;\xi^k\right)\right]\right\}\le$$

$$\le\sum_{k=1}^{t}\gamma_k\left(x^{k-1}-x\right)^T\nabla_x E_{\xi,x}\left[f_k\left(x^{k-1};\xi^k\right)\right]\le\beta_t V\left(x\right)-$$

$$-\sum_{k=1}^{t}\gamma_k\left(x^{k-1}-x\right)^T\left(\nabla_x f_k\left(x^{k-1};\xi^k\right)-\nabla_x E_{\xi,x}\left[f_k\left(x^{k-1};\xi^k\right)\right]\right)+ \qquad (¤)$$

$$+\sum_{k=1}^{t}\frac{\gamma_k^2}{2\beta_{k-1}}\left\|\nabla_x f_k\left(x^{k-1};\xi^k\right)\right\|_\infty^2.$$

Приложения

Для того чтобы применять МЗС к **многоруким бандитам**:

$$f_k\left(x;\xi^k\right)=r_i^k \text{ с вероятностью } x_i,\ i=1,...,n;$$

$$\text{а чтобы } \nabla_x E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right]=a^k=E\left[r^k\right], \text{ выбирают}$$

$$\nabla_x f_k\left(x;\xi_k\right)=(0,...,\underset{i}{r_i^k}/x_i,...,0)^T \text{ с вероятностью } x_i,\ i=1,...,n,$$

где r_i^k – потери (*regret*), которые “выдает” i-я ручка, если её дернуть на шаге k. При этом в (¤) вместо слагаемых

$\frac{\gamma_k^2}{2\beta_{k-1}}\left\|\nabla_x f_k\left(x^{k-1};\xi^k\right)\right\|_\infty^2$ нужно писать точнее $\gamma_k^2\frac{x_j^k\left(1-x_j^k\right)}{\beta_{k-1}}\left(\frac{r_j^k}{x_j^k}\right)^2$,

где j – номер ручки, выбранной алгоритмом на k-м шаге.

Для задачи **рандомизированного взвешивания экспертных решений**

$$\boxed{E_{\xi^k}\left[f_k\left(x;\xi^k\right)\right] = \sum_{i=1}^{n} \lambda\left(\omega^k, \zeta_i^k\right) x_i},$$

где $\lambda\left(\omega^k, \zeta_i^k\right)$ – потери Эксперта i, выбравшего на шаге k стратегию ζ_i^k, при ходе “сопротивляющейся Природы” ω^k.

Напомним схему игры: на каждом шаге первым выбирают свои ходы Эксперты, потом мы, потом природа; но есть важный нюанс: наш ход заключается в выборе распределения вероятностей, которое становится известным Природе, но разыгрывание согласно этому распределению вероятностей происходит уже после того, как Природа выбрала ω^k, то есть реализация Природе не известна.

Тонкая разница в постановке этих двух задача (“сто́ящая” $\sqrt{n}$ в оценке $M = \tilde{M}\sqrt{n}$ для многоруких бандитов), заключается в том, что в многоруких бандитах мы имеем только свою историю дергания ручек (нам не известно, какой бы *regret* принесли нам другие ручки, кабы мы их выбрал), а в постановке взвешивания экспертных решений это все известно и называется потерями экспертов.

В детерминированных задачах **взвешивания экспертных решений**:

$$\boxed{f_k\left(x;\xi^k\right) \equiv f_k\left(x\right) = \left\langle l^k, x\right\rangle}$$

для **задачи оптимального распределения доллара**,

а для **задачи из замечания 2** (детерминировано-выпуклый вариант задачи о взвешивания экспертных решений):

$$x := \sum_{i=1}^{n} x_i \cdot \zeta_i^k \overset{def}{=} \gamma,\ f_k\left(x;\xi^k\right) \equiv f_k(x) = \sum_{i=1}^{n} x_i \lambda\left(\omega^k, \zeta_i^k\right) \geq \lambda\left(\omega^k, \gamma\right)$$

(поскольку $\lambda\left(\omega^k, \zeta\right)$ – выпуклая по ζ для любого ω^k) $\Rightarrow$

$$\sum_{k=1}^{N} \lambda\left(\omega^k, \gamma^{k-1}\right) - \min_{i=1,\ldots,n} \sum_{k=1}^{N} \lambda\left(\omega^k, \zeta_i^k\right) \leq \sum_{k=1}^{N} f_k\left(x^{k-1}\right) - \min_{x \in S_n(1)} \sum_{k=1}^{N} f_k(x).$$

Неулучшаемость оценок МЗС

Удивительно, что во всех случаях, описанный МЗС дает (с мультипликативной точностью до константы) не улучшаемые оценки (в случае многоруких бандитов с мультипликативной точностью $\sim \sqrt{\ln n}$).

Алгоритм МЗС-разреженный

В случае решения детерминированных разреженных задач выпуклой оптимизации в пространствах огромных размеров, как, например, задачи о поиске PageRank:

$$f(x) = \max_{u \in S_n(1)} \langle u, (P^T - I)x \rangle \to \min_{x \in S_n(1)}$$

важную роль играет искусственное введение случайности с помощью рандомизации Григориадиса–Хачияна (1995). При этом сохраняются все приведенные выше оценки, но, к сожалению, теряется адаптивность, то есть теперь мы должны знать заранее либо требуемую точность, либо число шагов. Более того, оценки не только сохраняются, но и в $\sim \sqrt{2}$ раз становятся лучше за счет потери адаптивности.

Согласно распределению вероятностей

$$p_i^t = \frac{\exp\left(-\frac{1}{\beta_t}\sum_{k=1}^{t}\gamma_k \frac{\partial f_k\left(x^{k-1}\right)}{\partial x_i}\right)}{\sum_{l=1}^{n}\exp\left(-\frac{1}{\beta_t}\sum_{k=1}^{t}\gamma_k \frac{\partial f_k\left(x^{k-1}\right)}{\partial x_l}\right)}, \quad i = 1,\ldots,n,$$

получаем случайную величину $i(t)$,

$$x_{i(t)}^t = x_{i(t)}^{t-1} + 1, \;\; x_j^t = x_j^{t-1}, \; j \neq i(t).$$

Для этого алгоритма γ_t *и* β_t *разумнее брать постоянными:* $\gamma_t \equiv \beta_0^{-1}\sqrt{2/N} = M^{-1}\sqrt{2\ln n/N}$, $\beta_t \equiv 1$.

Литература

1. *Andersen S.P.*, *de Palma A.*, *Thisse J.-F.* Discrete choice theory of product differentiation. MIT Press, Cambridge, 1992.
2. *Юдицкий А.Б., Назин А.В., Цыбаков А.Б., Ваятис Н.* Рекуррентное агрегирование оценок методом зеркального спуска с усреднением // Проблемы передачи информации, 2005. Т. 41:4. С. 78–96.
3. *Lugoshi G.*, *Cesa-Bianchi N.* Prediction, learning and games. New York: Cambridge University Press, 2006.
4. *Juditsky A.*, *Nazin A.V.*, *Tsybakov A.B.*, *Vayatis N.* Gap-free Bounds for Stochastic Multi-Armed Bandit // IFAC World congress, 2008.

5. *Хачиян Л.Г.* Избранные труды / сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.
6. *Nesterov Y.* Primal-dual subgradient methods for convex problems // Math. Program. Ser. B. 2009. V. 120.
7. *Juditsky A.*, *Lan G.*, *Nemirovski A.*, *Shapiro A.* Stochastic approximation approach to stochastic programming // SIAM Journal on Optimization. 2009. V. 19. № 4. P. 1574–1609.
8. *Вьюгин В.В.* Элементы математической теории машинного обучения. М.: МФТИ, 2010, Глава 3.
9. *Devolder O.* Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. PhD Thesis, March 2013. CORE UCL. 309 p.
10. *Гасников А.В., Дмитриев Д.Ю., Нестеров Е.Ю.* О задаче ранжирования web-страниц PageRank // АиТ ???