Υπολογιστική Θεωρία Μηχανικής Μάθησης

Θέματα:

- Πιθανότητες
- Εκμάθηση κατανομών + Testing — Generative AI
- PAC Learning
  - Εκμάθηση διαχωριστών (classification)
  - Regression - Παλινδρόμηση
  - Επίδραση των σφαλμάτων + θορύβου στην μηχανική μάθηση
- Convex Optimization / Gradient Descent
- Online Convex Optimization / Multi-armed Bandits

- Δομή μαθήματος
    - Ασκήσεις (3-4 σειρές)
    - Παρουσίαση ενός paper
        - STOC FOCS
        - COLT
        - NeurIPS, ICML

Machine Learning

Statistics

Optimization

Algorithms

Νομική Ηθική

Ένα βασικό πρόβλημα

− Εκτίμηση της πιθανότητας ενός σφάλματος p

$$x_i \begin{cases} \xrightarrow{p} 1 & \text{κορώνα} \\ \xrightarrow{1-p} 0 & \text{γράμματα} \end{cases}$$

- Άμεση εφαρμογή : Δημοσκόπηση

Ένας απλός εκτιμητής είναι

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Γραμμικότητα της αναμενόμενης τιμής

$$E[\hat{p}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n} n \cdot p = p$$

- Αναμενόμενο σφάλμα

$$E[|\hat{p} - p|]$$

$$E[|\hat{p}-p|] \le \sqrt{E[(\hat{p}-p)^2]}$$

$$\le \sqrt{Var(\hat{p})}$$

$$Var(\hat{p}) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

επειδή X_i ανεξάρτητα

$$= \frac{1}{n^2} \sum_{i=1} Var(X_i)$$

$$= \frac{n}{n^2} \left(p(1-p)^2 + (1-p)p^2 \right)$$

$$= \frac{1}{n} p(1-p)$$

άρα $$E[|\hat{p}-p|] \le \sqrt{\frac{p(1-p)}{n}} \le \frac{1}{2\sqrt{n}}$$

$$\Rightarrow \quad E[\,|\hat{p}-p|\,] \le \varepsilon$$

με $\quad n \simeq \dfrac{1}{\varepsilon^2}$

Άνω φράγμα

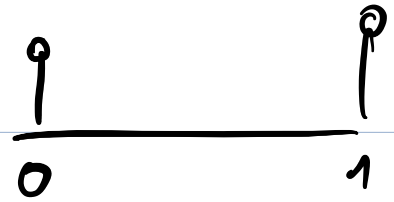
→ Ισχυρότερο Αποτέλεσμα (με μεγάλη πιθανότητα)

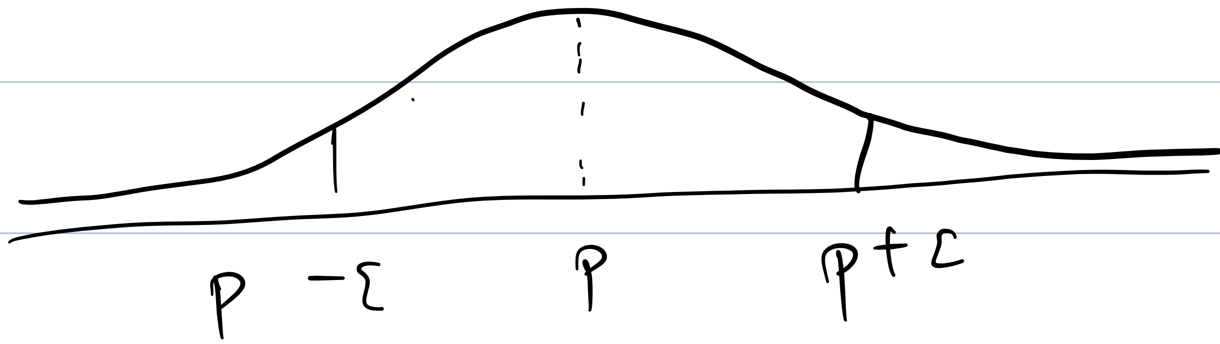
Πόσα δείγματα χρειάζονται ώστε

$$\Pr[\,|\hat{p}-p| > \varepsilon\,] < \delta$$

$1-\varepsilon$ $\quad \varepsilon$

$|\hat{p} - p|$





Ανισότητες Συγκέντρωσης

- Ανισότητα Markov

Αν X είναι μια μη αρνητική τυχαία μεταβλητή, τότε για κάθε $t>0$

$$\Pr[X \geq t] \leq \frac{E[X]}{t}$$

Εφαρμογή

Αν $X = |\hat{p} - p|$ τότε $X \geq 0$

$$\Pr[X \geq t] \leq \frac{E[X]}{t} \leq \frac{\frac{1}{2\sqrt{n}}}{t}$$

$$\Pr[X \geq \varepsilon] \leq \frac{1}{2\varepsilon\sqrt{n}}$$

$$\Pr[X > \varepsilon] \leq \delta$$ με

$$n = O\left(\frac{1}{\varepsilon^2\delta^2}\right)$$

Απόδειξη της Markov

$$1\{x \geq t\} = \begin{cases} 1 & x \geq t \\ 0 & \text{o/w} \end{cases}$$

$$E[X] = E\left[X\,1\{X \geq t\} + X\,1\{X < t\}\right]$$

$$= E[X\,1\{X \geq t\}]$$

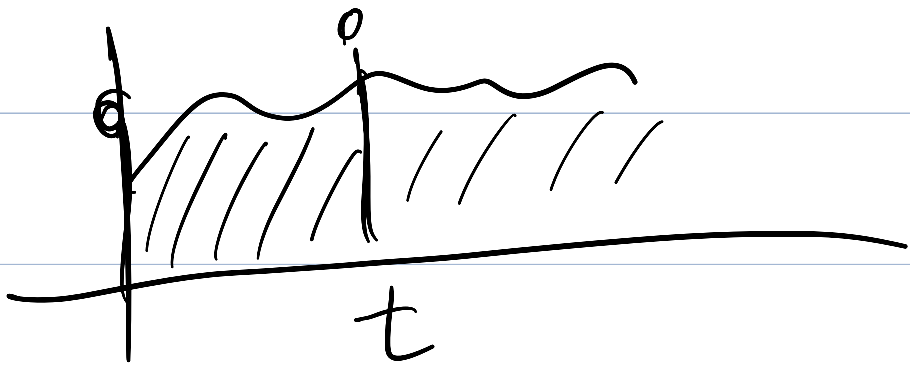
$$+ E[x \, 1\{x < t\}]$$

$$\geq E[x \, 1\{x \geq t\}]$$

$$\geq E[t \, 1\{x \geq t\}]$$

$$= t \Pr[x \geq t]$$

$$\Rightarrow \quad \Pr[x \geq t] \leq \frac{E[x]}{t} \qquad \square$$



Ανισότητα Chebyshev

Για κάθε τυχαία μεταβλητή X

με $Var[X] = \sigma^2$

$$\Pr[\,|X - E[X]| \geq k\sigma\,] \leq \frac{1}{k^2}$$

Απόδειξη

Θέτω $Z = (X - E[X])^2$

Ισχύει απο Markov για την

Z

$$\Pr[Z \geq t] \leq \frac{E[Z]}{t}$$

$$\Leftrightarrow \Pr[(X - E[X])^2 \geq t] \leq \frac{\sigma^2}{t}$$

για $k = \frac{\sqrt{t}}{\sigma}$

$$\Pr\left[(x - E[x])^2 \geq k^2 \sigma^2 \right] \leq \frac{1}{k^2}$$

$$\Leftrightarrow \Pr\left[|x - E[x]| \geq k\sigma \right]$$

$\square$

Chernoff Bound

Αν η τυχαία μεταβλητή $X = \sum_i X_i$ όπου X_i είναι ανεξάρτητες τυχαίες μεταβλητές με $X_i \in \{0,1\}$ and $E[X] = \mu$

$$\Pr\left[X \geq (1+\varepsilon)\mu \right] \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}} \right)^{\mu}$$

$$\leq e^{-\frac{\varepsilon^2 \mu}{2+\varepsilon}}$$

Απόδειξη

$$Z = e^{tX} \quad \text{όπου } t>0$$

από Markov για τη Z

$$\Pr[e^{tX} \geq e^{t(1+\varepsilon)\mu}]$$

$$\leq \frac{E[e^{tX}]}{e^{t(1+\varepsilon)\mu}} \leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\varepsilon)\mu}}$$

$$E[e^{tX}] = E[e^{t\sum X_i}]$$

$$= E[\prod_i e^{tX_i}]$$

$$= \prod_i E[e^{tX_i}]$$

$$= \prod_i ((1-p_i)e^0 + p_i e^t)$$

$$= \prod_i (1 + p_i(e^t-1))$$

$$(1+x \leq e^x) \quad \leq \prod_i e^{p_i(e^t-1)}$$

$$= e^{\sum p_i (e^t-1)} = e^{\mu(e^t-1)}$$

Για $t = \log(1+c)$ ✓ $\square$

Εφαρμογή για την εκτίμηση του κέρματος

$$E\left[\sum X_i\right] = np$$

Το Chernoff bound μας δίνει

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1+q)\,np\right] \leq e^{-\frac{q^2 np}{2+q}}$$

Θέτοντας $q = \dfrac{\varepsilon}{p}$

$\hat{p}$

$$\Pr\left[\overbrace{\frac{1}{n}\sum_{i=1}^{n} X_i} \geq p + \varepsilon\right]$$

$$\leq e^{-\frac{\varepsilon^2}{2p+\varepsilon} n} \leq \frac{\delta}{2}$$

άν

$$n = O\left(\frac{p+\varepsilon}{\varepsilon^2} \log(1/\delta)\right)$$

$$\Pr[\hat{p} \geq p+\varepsilon] \leq \delta/2$$

Με ένα συμμετρικό επιχείρημα

$$\Pr[\hat{p} \leq p-\varepsilon] \leq \delta/2$$

$\Pr[|\hat{p}-p| \geq \varepsilon] \leq$ (union

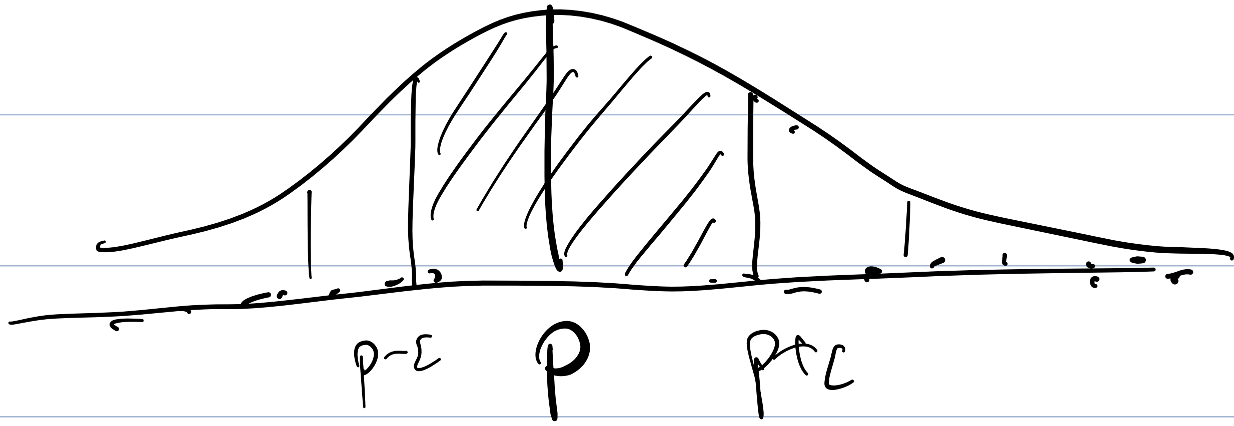
Βολλη)

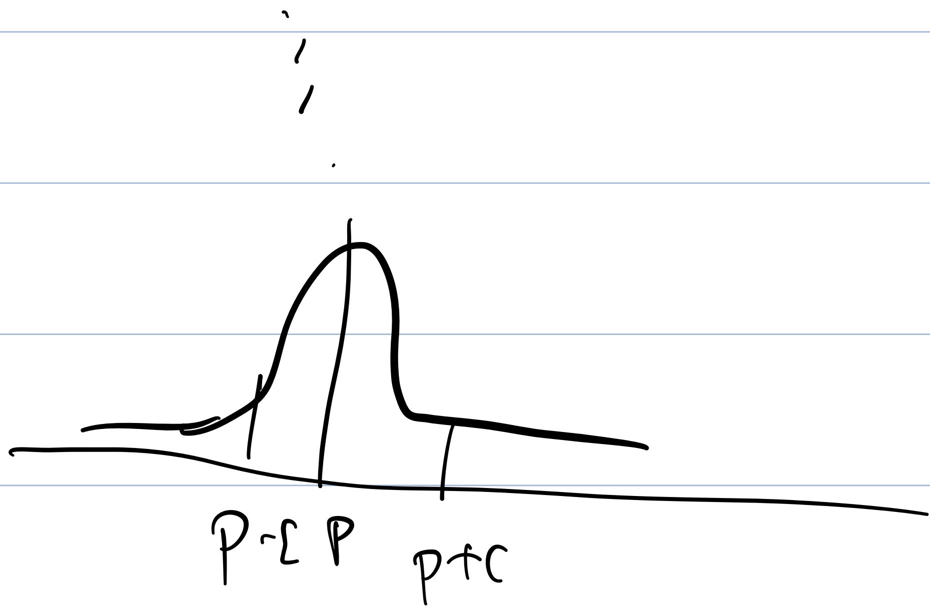
$$\Pr[\hat{p} - p \geq \varepsilon] + \Pr[\hat{p} - p \leq -\varepsilon] \leq \delta/2 + \delta/2 \leq \delta$$

$\boxed{\text{Πόρισμα: Αν } n = O\left(\frac{1}{\varepsilon^2}\log(1/\delta)\right) \text{ δείγματα βρίσκω } \hat{p} \text{ τέτοιο ώστε } |\hat{p} - p| \leq \varepsilon \text{ με πιθανότητα } 1-\delta}$

Ακρίβεια ε

Πιθανότητα σφάλματος δ


p-ε
p
p+ε


p-ε
p
p+ε

Ανισότητα Hoeffding

Αν $X = \sum X_i$ με $a_i \le X_i \le b_i$

και X_i είναι ανεξάρτητα

Τότε

$$\Pr\left[|X - E[X]| \ge t \right] \le 2e^{-\frac{2t^2}{\sum (b_i - a_i)^2}}$$

$$\Pr\left\{ |\hat{p} - p| \ge \varepsilon \right\} \le 2e^{-\frac{2\varepsilon^2 n^2}{n}} = 2e^{-2\varepsilon^2 n}$$

$$n = O\left(\frac{1}{\varepsilon^2} \log(1/\delta) \right)$$