3/1/19

EE5603: Concentration Inequalities

- Happy New Year 2019!
- Google classroom code: $\boxed{\text{bsazdx}}$
- Today:
  - Law of large numbers (LLN)
  - Central limit theorem
  - From asymptotic analysis to finite sets
  - Connection with machine learning
  - Review: Markov inequality, Chebyshev inequality

LLN: If $X_1, \ldots X_n$ are i.i.d observations of a random variable with mean μ, and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

WLLN: $\bar{X}_n \xrightarrow{p} \mu$

SLLN: $\bar{X}_n \xrightarrow{a.s.} \mu$

Observation: Asymptotic result

Central limit theorem: If $X_1, \ldots X_n$ is a sequence of iid RVs with mean μ and finite variance σ^2, with

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Observation: Asymptotic result

Aside: $a.s. \Rightarrow p \Rightarrow d$

$m.s \Rightarrow p$

If $X_1, \ldots X_n$ is a seq of RVs (i.i.d), then

- $X_n \xrightarrow{p} X$ when $P(\lim_{n\to\infty} |X_n - X| \geq \varepsilon) = 0$ for any $\varepsilon > 0$
- $X_n \xrightarrow{a.s.} X$ when $P(\lim_{n\to\infty} X_n = X) = 1$
- $X_n \xrightarrow{d} X$ when $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$
- $X_n \xrightarrow{m.s} X$ when $\lim_{n\to\infty} E[(X_n - X)^2] = 0$

<u>Supervised learning</u> (SL): Let $(x_1, y_1) \ldots, (x_n, y_n)$ be n training samples coming from a joint distribution $p(x,y)$. Let $L(x,y)$ be our loss function and $f(x;\alpha)$ be our machine; then

expected risk : $R(\alpha) = \int L(y, f(x;\alpha))\, dP(x,y)$

The goal of SL is to find α_0 that minimizes the risk functional $R(\alpha)$.. This is not achievable since $p(x,y)$ is not known in practise.

$\therefore$ Empirical risk: $R_{emp}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i;\alpha))$

Recall: • Markov inequality: For a non-negative RV X, and for any $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon}$$

• Chebyshev inequality: For a random variable X, for any $\varepsilon > 0$

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

• Chernoff bound: For a random variable X, for any $t > 0$,

$$P(e^{Xt} \geq e^{\varepsilon t}) \leq \frac{E[e^{Xt}]}{e^{at}}$$