Lecture 4

Recap:

→ Defined Agnostic PAC

→ Defined Uniform Convergence

→ Uniform Convergence for finite hypothesis classes.

→ Concentration inequalities

Let $\{x_1, x_2, \ldots\}$ be sequence of random variables $\mathbb{E}[x_i] = \mu$, $Var(x_i) = \sigma^2 < \infty$. Let $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$. How close is $\bar{x}_n$ to μ?

Today:

→ Finish Concentration Inequality

→ VC dimension

Chernoff-style bounds.

$$\log\left(\Pr[X - \mathbb{E}(X) \geq t]\right) \leq \inf_{\lambda \geq 0}\left(\log\left(\mathbb{E}\left(e^{\lambda(X-\mathbb{E}(X))}\right)\right) - \lambda t\right).$$

Gaussian tail bounds

$X \sim N(\mu, \sigma^2)$

Goal: Bound $\Pr[X - \mu \geq t]$

$$\Pr[X - \mu \geq t] \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}\left(e^{\lambda(X-\mathbb{E}(X))}\right)}{e^{\lambda t}}$$

$X - \mathbb{E}[X] = y$

$y \sim N(0, \sigma^2)$

$$\mathbb{E}\left(e^{\lambda(X-\mathbb{E}(X))}\right) = \mathbb{E}\left(e^{\lambda y}\right)$$

$$= \int_{\mathbb{R}} e^{\lambda y} \frac{e^{-y^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}\, dy$$

$$= \int_{\mathbb{R}} \frac{e^{-\left(y-\frac{\lambda}{\sigma^2}\right)^2 \cdot \frac{1}{2\sigma^2}}\, e^{\frac{\lambda^2\sigma^2}{2}}}{\sqrt{2\pi}\,\sigma}\, dy$$

$$= e^{\lambda^2\sigma^2/2} \int_{\mathbb{R}} \frac{e^{-\left(y-\frac{\lambda}{\sigma^2}\right)^2 \cdot \frac{1}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} dy$$

$$= e^{\frac{\lambda^2\sigma^2}{2}}$$

$$\log\left(\Pr[X - \mathbb{E}[X] \geq t]\right) \leq \inf_{\lambda \geq 0} \left(\frac{\lambda^2\sigma^2}{2} - \lambda t\right)$$

$$= \frac{-t^2}{2\sigma^2} \qquad \text{[Exercise]}$$

$$\Pr[X - \mathbb{E}[X] \geq t] \leq e^{-t^2/2\sigma^2}$$

$$\leq e^{-t^2/2\sigma^2} \cdot (\mathrm{poly}(t,\sigma))^{-1}$$

(Tight upto polynomial factors, best bound one can show is above)

If $\overline{X}_n = \sum_{i=1}^{n} X_i / n \qquad \mathbb{E}[X_i] = \mu \qquad \mathrm{Var}[X_i] = \sigma^2$

Gaussian is a stable distribution

$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$

$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$\overline{X}_n \sim N(\mu, \sigma^2/n)$

$$\therefore \Pr[\overline{X}_n - \mu \geq t] \leq e^{-nt^2/2\sigma^2}$$

$\forall\, n, \delta$, $\quad \Pr\left[\overline{X}_n - \mu \geq \sigma\sqrt{\frac{2\log(1/\delta)}{n}}\right] \leq \delta.$

Sub-Gaussian random variable.

A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian if there exists a positive numbers σ such that

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le e^{\sigma^2\lambda^2/2} \quad \forall \lambda \in \mathbb{R}.$$

σ : sub-Gaussian parameter (variance proxy)

Any Gaussian random variable with σ^2 is sub-Gaussian with parameter σ.

Many non-Gaussian random variables also have this property!

Rademacher r.v. : $X : \{-1, +1\}$ with equal probability

Claim : sub-Gaussian with parameter $\sigma = 1$.

Proof . $\mathbb{E}[X] = 0$

$$\mathbb{E}\left[e^{\lambda X}\right] = \frac{1}{2}\left(e^{-\lambda} + e^{\lambda}\right)$$

$$= \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty} \frac{(\lambda)^k}{k!}\right)$$

$$= \sum_{k=0}^{\infty} \frac{(\lambda)^{2k}}{(2k)!}$$

$$\leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{2^k k!} \qquad \left((2k)! \geq 2^k k!\right)$$

$$= e^{\lambda^2/2}$$

■

Concentration bound for sum of sub-Gaussian r.v.

Theorem: Suppose that the random variables $\{X_i,\ i = \{1, \dots n\}\}$ are independent and $\mathbb{E}[X_i] = \mu_i$ and X_i is sub-Gaussian with parameter σ_i. Then for all $t \geq 0$, we have

$$\Pr\left[\left|\sum_{i=1}^{n} (X_i - \mu_i)\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} \sigma_i^2}\right).$$

Proof.

We show that $Z = \sum_{i=1}^{n} X_i$ is sub-Gaussian with parameter $\sigma_Z^2 = \sum_{i=1}^{n} \sigma_i^2$.

$$\mathbb{E}\left(e^{\lambda(Z - \mathbb{E}(Z))}\right) = \mathbb{E}\left(e^{\lambda(\sum X_i - \mathbb{E}(\sum X_i))}\right)$$

$$= \prod_{i=1}^{n} \mathbb{E}\left(e^{\lambda(X_i - \mathbb{E}(X_i))}\right) \qquad \text{(independence)}$$

$$\leq \prod_{i=1}^{n} e^{\sigma_i^2 \lambda^2/2} = e^{(\sum \sigma_i^2)\lambda^2/2}$$

The Theorem now follows from the following Lemma.

Concentration bound for sub-Gaussian r.v.

Lemma: If X is sub-Gaussian with parameter σ, then for any $t > 0$,

(1) $\Pr[X > \mathbb{E}(X) + t] \leq e^{-t^2/2\sigma^2}$.

(2) $\Pr[X < \mathbb{E}(X) - t] \leq e^{-t^2/2\sigma^2}$.

(3) $\Pr[|X - \mathbb{E}(X)| \geq t] \leq 2e^{-t^2/2\sigma^2}$.

Proof of Lemma.

(1)
$$\log(\Pr[X - \mathbb{E}(X) \geq t]) \leq \inf_{\lambda \geq 0} \left(\log \mathbb{E}\left(e^{\lambda(X - \mathbb{E}(X))}\right) - \lambda t \right)$$

By sub-Gaussianity,

$$\mathbb{E}\left(e^{\lambda(X - \mathbb{E}(X))}\right) \leq e^{\sigma^2 \lambda^2/2}$$

$$\log(\Pr[X - \mathbb{E}(X) \geq t]) \leq \inf_{\lambda \geq 0} \left(\frac{\sigma^2 \lambda^2}{2} - \lambda t \right) = -\frac{t^2}{2\sigma^2}$$

(1) follows.

(2) Take $X' = -X$.

X' is also sub-Gaussian with parameter σ

(by symmetry of the definition wrt λ)

$$\Pr[X < \mathbb{E}(X) - t] = \Pr[X' > \mathbb{E}(X') + t].$$

Result follows by (1).

(3) Union bound

$\blacksquare$

Corollary of Thm for sum of sub-Gaussian

Consider n i.i.d. r.v. $X_1, \ldots, X_n$ each with $\mathbb{E}(X_i) = \mu$ & sub-Gaussian parameter σ.

Then for all $t \geq 0$

$$\Pr\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \geq \varepsilon\right] \leq \exp\left(\frac{-n\varepsilon^2}{2\sigma^2}\right).$$

Bounded random variable is sub-Gaussian

X is zero-mean supported on interval $[a, b]$.

Claim. X is sub-Gaussian with parameter $b-a$.

(Possible to show sub-Gaussian with parameter $\frac{b-a}{2}$.)

Proof 1

$\sigma = b - a.$

$$\mathbb{E}(e^{\lambda X}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!}\right)$$

For any k,

$\mathbb{E}(X^k) \le \mathbb{E}(|X|^k)$

$\le (b-a)^k$

$= \sigma^k$

$$= \sum_{k=0}^{\infty} \lambda^k \frac{\mathbb{E}(X)^k}{k!}$$

$$\le \sum_{k=0}^{\infty} \frac{\lambda^{2k} \cdot \sigma^{2k}}{(2k)!}$$

(Error!)

Not true

$\mathbb{E}[X] = 0$

but $\mathbb{E}[X^3]$

could be > 0

$$\le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k} \sigma^{2k}}{2^k k!}$$

$$= e^{\lambda^2 \sigma^2 / 2}$$

■

Thank you for finding a bug here today!

Here's one fix which gets a slightly worse constant:

$$\mathbb{E}(e^{\lambda x}) = \sum_{k=0}^{\infty} \lambda^k \frac{\mathbb{E}(x^k)}{k!}$$

$$\mathbb{E}(x^k) \le \mathbb{E}(|x|^k) \le \sigma^k$$

Also, $\mathbb{E}(x) = 0$

$$\therefore \mathbb{E}(e^{\lambda x}) \le 1 + \sum_{k \ge 2}^{\infty} \frac{\lambda^k \sigma^k}{k!}$$

$$= e^{\lambda\sigma} - \lambda\sigma$$

$$\le e^{\lambda^2 \sigma^2}$$

$$\left(\text{since } e^{x} - x \le e^{x^2} \ \forall x\right)$$

Therefore, we've shown sub-Gaussian with parameter $\sqrt{2}\sigma = \sqrt{2}(b-a)$.

The above proof has the advantage of being more analytic, but the proof below is cleverer and gets a better constant.

Proof 2

Symmetrization : Let x' be an independent copy of x.

$\mathbb{E}[x'] = 0.$

$$\mathbb{E}_x[e^{\lambda x}] = \mathbb{E}_x\left[\exp\left(\lambda\left(x - \mathbb{E}(x')\right)\right)\right]$$

$$\leq \mathbb{E}_{x,x'}\left(\exp\left(\lambda\left(x - x'\right)\right)\right)$$

(Jensen's inequality : $f(\mathbb{E}(x)) \leq \mathbb{E}(f(x))$ for any convex f)

$(\mathbb{E}(x))^2 \leq \mathbb{E}(x^2)$

$$= \mathbb{E}_{x,x',\varepsilon}\left(\exp\left(\lambda \varepsilon\left(x - x'\right)\right)\right)$$

ε : Rademacher r.v. $\{\pm 1\}$ with prob. $1/2$

$x - x'$ has same dist as $x' - x$

$$\mathbb{E}[e^{\lambda x}] \leq \mathbb{E}_{x,x'}\left(\exp\left(\lambda^2 (x - x')^2/2\right)\right)$$

Implies

$$\leq \exp\left(\lambda^2 (b-a)^2/2\right) \quad \blacksquare$$

$$\underline{\mathbb{E}_{x,x'}\left(\exp\left(\lambda\left(x - x'\right)\right)\right) = \mathbb{E}_{x,x'}\left(\exp\left(\lambda\left(x' - x\right)\right)\right)}$$

Lemma (weaker version of Hoeffding's)

Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$ for each $i \in [n]$.
Then for any $\epsilon > 0$,

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]\right| \leq \varepsilon\right]$$

$$\geq 1 - 2\exp\left(-\frac{n^2\varepsilon^2}{2\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Remark: with right sub-Gaussianity, we'll get right bound.

Thm. (McDiarmid's Inequality)

Let $x_1, \ldots, x_n$ be independent r.v. taking values over some domain $\mathcal{X}$. Let $f: \mathcal{X}^n \to \mathbb{R}$ which satisfies the bounded differences property:

$\forall\ i \in [n]$ & $\forall\ x_1, \ldots x_n, x_i'$ ($x_i' \in \mathcal{X}$).

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots x_n)$$
$$- f(x_1 \ldots, x_{i-1}, x_i', x_{i+1}, \ldots x_n)| \leq c_i.$$

If f satisfies this property, then

$$\Pr\left[f(x_1, \ldots, x_n) - \mathbb{E}(f(x_1 \ldots x_n)) > \varepsilon\right]$$
$$\leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n} c_i^2}\right).$$

Remark: Gives Hoeffdings $\left(f = \left(\frac{\sum x_i}{n}\right)\right.$

$$\left. c_i = \frac{b_i - a_i}{n} \quad \forall i\right)$$