Least Squares is MLE under Gaussian Noise

Let $x_0 \in \mathbb{R}^n$

Let $a_i \in \mathbb{R}^n \quad i=1\dots m$

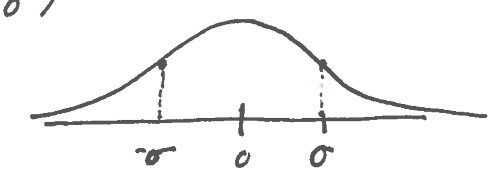
Let $\varepsilon_i \sim N(0, \sigma^2) \quad i=1\dots m$

Let $b_i = a_i \cdot x_0 + \varepsilon_i$

Given $\{a_i, b_i\}_{i=1}^{m}$ Find x_0.

Recall density function for $N(\mu, \sigma^2)$

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Likelihood function of b is ~~{for known σ}~~

$$L(x) = \frac{1}{(2\pi)^{\frac{m}{2}}\sigma^m} e^{-\frac{\sum_{i=1}^{m}(b_i - a_i \cdot x)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\varepsilon_1^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\varepsilon_2^2}{2\sigma^2}} \times \dots \times \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\varepsilon_n^2}{2\sigma^2}}$$

$$\arg\max_x L(x) = \arg\max_x \; -\sum_{i=1}^{m} \frac{(b_i - a_i \cdot x)^2}{2\sigma^2} = \arg\min \sum_{i=1}^{m} (b_i - a_i \cdot x)^2$$

So maximizing likelihood is equivalent to minimizing least squares error (under Gaussian noise).

Activity:

Let $X_i \sim N(0,\sigma)$ $i = 1 \dots \infty$

Plot what $\max\limits_{i=1\dots n} |X_i|$ will look like?

How do you think it scales w/ large n?

$\mathbb{P}(|N(0,\sigma)| \geq t)$ for large t?

How long on avg until you see something at least t?

How big after N samples?

Max of many Gaussians

Let X_i be r.v. st $\mathbb{E}(e^{tX_i}) \le e^{t^2\sigma^2/2}$ $\forall$ $t \ge 0$,

Then $\mathbb{E}(\max_{i\in[n]} X_i) \le \sigma\sqrt{2\lg n}$

Pf: By Jensen's inequality

$$e^{t\,\mathbb{E}(\max X_i)} \le \mathbb{E}(e^{t\max X_i})$$

$$= \mathbb{E}(\max e^{tX_i}) \le \sum_{i=1}^{n} \mathbb{E}(e^{tX_i}) \le n e^{t^2\sigma^2/2}$$

Thus, $\mathbb{E}(\max X_i) \le \frac{\lg n}{t} + \frac{t\sigma^2}{2}$

set $t = \sqrt{2\lg n}/\sigma$ to get $\mathbb{E}(\max X_i) \le \sigma\sqrt{2\lg n}$

Note: For $X_i \sim N(0,\sigma^2)$, $\mathbb{E}(e^{tX_i}) = e^{t^2\sigma^2/2}$.

Probability Bounds for max of many Gaussians

Let $X_i \sim N(0,1)$ $i=1 \cdots n$

$$P\left(\max |X_i| > \sqrt{2(1+\gamma)\log n}\right) \leq \sqrt{\frac{2}{\pi}}\, n^{-\gamma}$$

Proof by union bound

Proof: $\mathbb{P}(\max |X_i| > k) \leq n\, P(|X_i| > k)$.

Recall $\mathbb{P}(|X_i| > k) \leq \sqrt{\frac{2}{\pi}} \frac{1}{k} e^{-k^2/2}$

So $\mathbb{P}(\max |X_i| > k) \leq \sqrt{\frac{2}{\pi}}\, n e^{-k^2/2}$ (for $k > 1$)

Plugging in $k = \sqrt{2(1+\gamma)\log n}$ we get

$$\mathbb{P}\left(\max |X_i| > \sqrt{2(1+\gamma)\log n}\right) \leq \sqrt{\frac{2}{\pi}}\, n e^{-(1+\gamma)\log n} = \sqrt{\frac{n}{\pi}}\, n^{-\gamma} \quad \blacksquare$$

~~To get a better~~.

Note: tradeoff between tightness & failure probability

smaller γ means tighter control on upper bound of $\max |X_i|$

but slower decay of probability