موضوعات منتخب در فیزیک آماری

Session 5:

1403 ,12,12

دسته بندی روش های Optimization

1) Gradient Methods

1-1) First order methods $-\nabla L(\vec{x})$

a) Gradient Descent (GD)
b) Stochatic " " (SGD)
c) Momentum based GD
d) Nesterov Accelerated
e) Adam
f) RMS prop

1-2) Second-order Methods (using Hesian second Der)

a) Neton's Method
b) Quasi - Newton Method — تقریب H
c) Guass - Newton Method

این ها برای توابع بسیار تیز که بد هستند استفاده می‌شوند => راه فراری از Local min ندارند.

برای ارائه بچه‌های درس موضوع Automation derivative

2) Non - Gradiant => Swarm intelligence / Nature ispired

2-1) PSO (particle swarm opt)

2-2) Ant Colony (ACO)

2-3) Artificial Bee colony (ABC)

2-4) Simulated Anneling

2-5) fire flight Alg

3) Stochastic and Evolution Alg

a) Genetic Alg (GA)

b) Differential Evolution (DE)

c) Evolution strategies

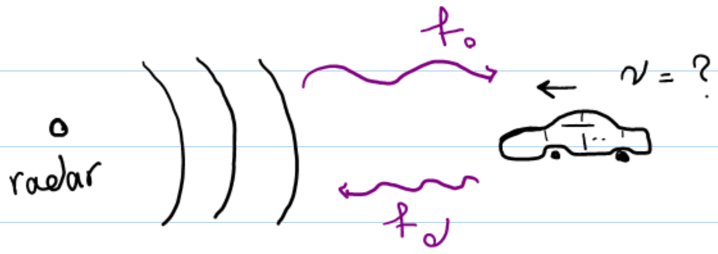
Introduction to parameter estimation:

$\vec{X}$ داده را داریم که آزمایش ها است. که توزیعی دارد.

$\vec{x}$ آزمایش $\longrightarrow f(\vec{x}, \vec{\alpha})$

پارامتر نامشخص

We want estimate α from observed $\vec{x}$.



سؤال)

$$x[n] = A\cos[2\pi f_d n + \phi] + w[n]$$

Noise

estimation

$$\hat{f}_d = g[x[0]\ x[1], \ldots, x[N-1]]$$

هر چه هر چه فرض می‌کنند می‌توانند بعید.

In general: $\hat{\alpha} = g(\vec{x})$ be the estimate of $\vec{\alpha}$

Questions?

- How do you find $g(\vec{x})$?
- Which $g(\vec{x})$ is good?

or

How good is $g(\vec{x})$?

$\vec{x}$ is random $\Rightarrow$ $\hat{\alpha}$ is random too

But we have a real α too: α_r

- Bias

$$b(\vec{\alpha}) = E(\hat{\vec{\alpha}}) - \vec{\alpha}_r$$

if $b(\vec{\alpha}) = 0 \Rightarrow \hat{\vec{\alpha}}$ is unbiased

- Covariance

$$\vec{C}(\hat{\vec{\alpha}}) = E\left[(\hat{\vec{\alpha}} - E[\hat{\vec{\alpha}}])(\hat{\vec{\alpha}} - E[\hat{\vec{\alpha}}])^T\right]$$

$$[C(\hat{\vec{\alpha}})]_{ii} = \text{variance}$$

3) Mean squared error:

$$MSE\{[\hat{\vec{\alpha}}]_i\} = E\{[\hat{\vec{\alpha}} - \vec{\alpha}_r]_i^2\}$$

$$= \vec{C}(\hat{\vec{\alpha}}) + b(\hat{\vec{\alpha}})\, b^T(\hat{\vec{\alpha}})$$

$$\Rightarrow \begin{cases} \vec{x} \longrightarrow f(\vec{x}, \alpha) \\ \hat{\alpha} \longrightarrow g(\vec{x}) \end{cases}$$

2 روش برای محک وجود دارد که g را تخمین بزنند و تشخیص بدهد کدام بهتر است.

I) Maximum likelihood

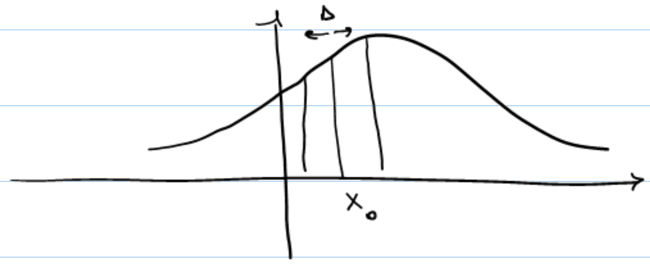
II) Baysian Method

(I) Maximum liklihood:

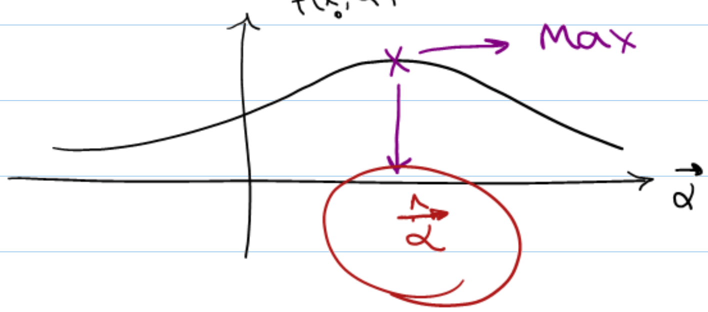
$$\hat{\vec{\alpha}} = \arg\max_{\vec{\alpha}} \{ f(x_0 , \alpha) \}$$

یک نقطه خاص

In 1-D



$$\int f(\vec{x}, \vec{\alpha}) \, dx = f(\vec{x}, \vec{\alpha}) \, \Delta$$



Why does this work?

هر آنچه بیشتر این روش نیستیم ، Var ای که به ما می‌رسد بزرگ‌تر خواهد بود.

این estimator کوچک‌ترین var(α) را دارد.

(II) Baysian Method:

تابع توزیع داریم احتمال آنکه a و b رخ دهند $\Leftarrow$ f(a, b)

شرطی آن جزوش داریم :

$$f(a,b) = f(a|b)\, f(b) = f(b|a)\, f(a)$$

$$\text{و} \quad \int f(a,b)\, da\, db = 1$$

حال α و x بذاریم :

احتمال α زمانی که x اندازه گیری شده

Posterior

آزمایش

$$f(\vec{\alpha}|\vec{x}) = \underbrace{f(\vec{x}|\vec{\alpha})}_{\text{likelihood}} \frac{f(\vec{\alpha})}{f(\vec{x})}$$

Prior

evidence ← data توزیع

پیشین ⇐ ایده داریم این پارامتر چقدر احتمال دارد ⇐ نوعی اطلاعات

مثال از Maximum likelihood :

Calling some number from computer $\sim \mathcal{N}(\mu, \sigma^2)$

Prior

$$n = 0, \dots, N-1$$

حال چند اندازه گیری می‌دهد ⇐ $x[0], \dots, x[N-1]$

مستقل هستن

می‌خواهیم μ و σ^2 را تخمین بزنیم

$$\Rightarrow f(x[0], \dots, x[N-1], \mu, \sigma^2)$$

$$= \frac{1}{(2\pi)^{N/2}} \frac{1}{(\sigma^2)^{N/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum (x[n]-\mu)^2 \right\}$$

خود تابع Max بشود log هم Max می‌شود پس Max log را می‌دهد.

$$\max_{\mu,\sigma^2} \left[-\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n]-\mu)^2 \right]$$

$$\begin{cases} \dfrac{\partial(\ )}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1} x[n] \\ \dfrac{\partial(\ )}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1} [x[n]-\hat{\mu}]^2 \end{cases}$$

مثال پواسون: توزیع ۲ جمله‌ای که p و یا q بسیار کوچک است.

واپاشی ماده radio active

در پواسون λ که همان نرخ هست را می‌خواهیم تخمین بزنیم

$$\underset{\text{ذره}}{f}[\alpha_i, \lambda] = \frac{\lambda^{\alpha_i}}{\alpha_i!} e^{-\lambda} \qquad \alpha_i \geq 0$$

مشاهده $\alpha_1, \alpha_2, \dots, \alpha_N$ ← مستقل ← تخمین random generator

می‌خواهیم λ را تخمین بزنیم

$$P(\alpha_1, \dots, \alpha_N, \lambda) = \prod_{n=1}^{N} \frac{e^{-\lambda} \lambda^{\alpha_n}}{\alpha_n!}$$

$$= C \, e^{-N\lambda} \lambda^{\sum_{n=1}^{N} \alpha_n}$$

$$\log \Rightarrow \quad -N\lambda + \left(\sum_{n=1}^{N} \alpha_n\right) \ln \lambda$$

$$\frac{\partial(\ )}{\partial \lambda} = 0 \Rightarrow \quad \hat{\lambda}_{MLE} = \frac{1}{N} \sum_{n=1}^{N} \alpha_n$$

در کتاب‌های قدیمی‌تر می‌توان دید:

$$MLE \Rightarrow \hat{\vec{\alpha}} = \underset{\vec{\alpha}}{\arg\min} \left(E\left[\underbrace{h}_{\text{cost function}}(\hat{\vec{\alpha}} - \vec{\alpha}) \right] \right)$$

اما می‌دانیم در این روش $Var(\hat{\vec{\alpha}})$ لزوماً min ترین نیست.

$$\text{typically} \quad h(\hat{\vec{\alpha}}, \vec{\alpha}) = \begin{cases} |\hat{\vec{\alpha}} - \vec{\alpha}|^2 \\ |\hat{\vec{\alpha}} - \vec{\alpha}| \end{cases}$$