Parameter Estimation

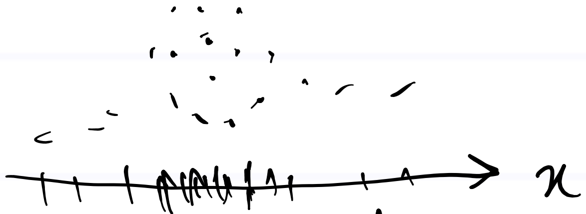
EE698V - Machine Learning for Signal Processing

Vipul Arora

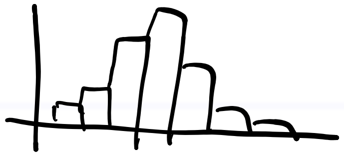
References

- PRML Section: 1.2 (**highly recommended**)

Recap



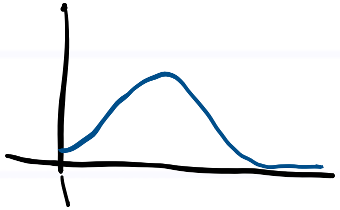
samples



histogram

inference

or



Samples

Model

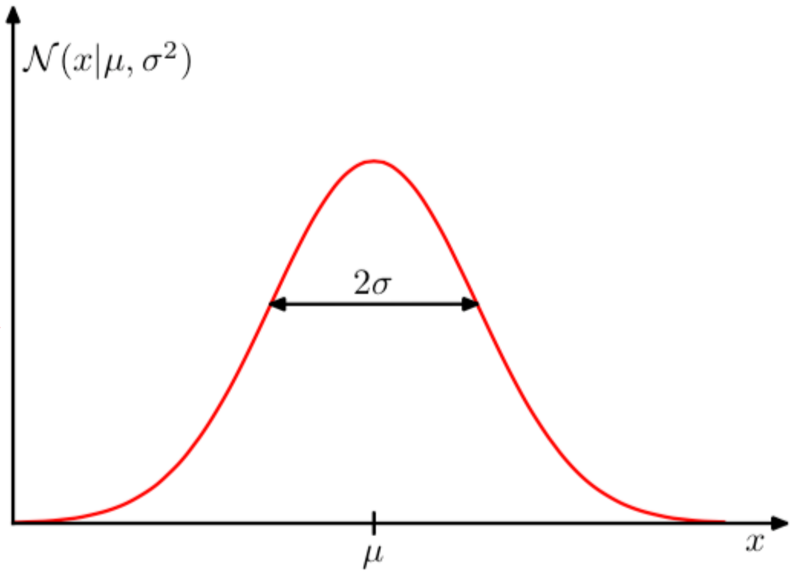
Standard distributions

- Known or easy-to-compute properties. E.g.,
  - integration
  - marginalization
  - derivatives
- Examples
  - Bernoulli, Multinomial, etc.
  - Gaussian, Laplacian, etc.

Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \underbrace{\frac{1}{(2\pi\sigma^2)^{1/2}}} \exp\left\{-\frac{1}{2\sigma^2}\underbrace{(x-\mu)^2}\right\}$$

$$e^{-(x-\mu)^2/2\sigma^2}$$



mean

$$\underline{\mathbb{E}[x]} = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \,\mathrm{d}x = \mu.$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2.$$

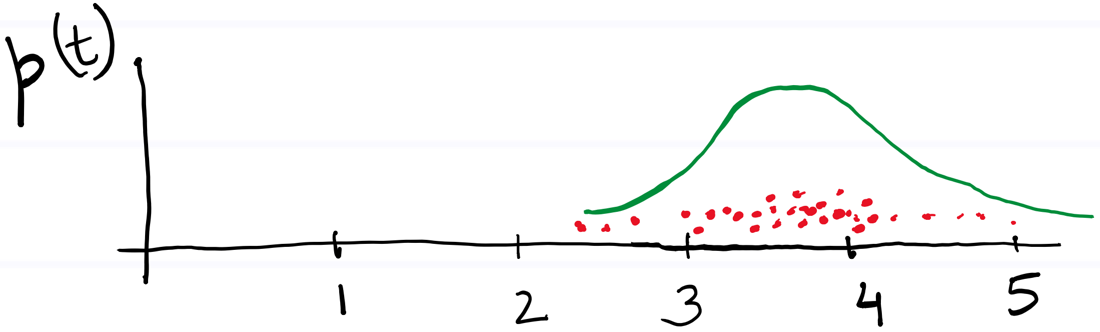
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x\, e^{-(x-\mu)^2/2\sigma^2} \, dx$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} \, dx = 1$$

Subs. $(x-\mu)^2 = \lambda$

Estimating the parameters

- I have samples of a r.v.. I want to model it with Gaussian distribution. How do I estimate the model parameters? μ, σ
- Random variable is t
- I want to maximize the probability of observed samples $\{s_1, s_2, \dots, s_N\}$



Estimating the parameters

- For N independent measurements
  - $p(t_1, t_2, \dots, t_N) = p(t_1)p(t_2)\dots p(t_N)$
- We want to estimate
  - $\boxed{\mu, \sigma = \underset{\mu,\sigma}{\operatorname{argmax}} \ln p(t_1, \dots, t_N | \mu, \sigma)\big|_{t_1=s_1,\dots,t_N=s_N}}$
- This is called maximum likelihood estimate (MLE)

MLE for parameters of Gaussian pdf

- $p(t_1, \ldots, t_N|\mu, \sigma) = \prod_{i=1}^{N} p(t_i|\mu, \sigma)$
- Products are difficult to handle, make it sum
- $\ln p(t_1, \ldots, t_N|\mu, \sigma) = \sum_{i=1}^{N} \ln p(t_i|\mu, \sigma)$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(t_i - \mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$$

$p(t_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t_i-\mu)^2/2\sigma^2}$

MLE for parameters of Gaussian pdf

- $\mathcal{L} = p(t_1, \dots, t_N | \mu, \sigma)|_{t_1 = s_1, \dots, t_N = s_N}$ is called likelihood
- $\ln \mathcal{L} = \ln p(t_1, \dots, t_N | \mu, \sigma)|_{t_1 = s_1, \dots, t_N = s_N}$ is called log likelihood
- Obtain μ, σ by maximizing the log likelihood ($\partial \mathcal{L} / \partial \mu = 0$ and $\partial \mathcal{L} / \partial \sigma^2 = 0$)

ML estimate of μ

- $\ln \mathcal{L} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(s_i - \mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$

- $\hat{\mu} = \text{argmax}_\mu \ln \mathcal{L}$

$$= \text{argmax}_\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(s_i - \mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$$

$$\frac{\partial(\ln \mathcal{L})}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_i (s_i - \mu)\,2(-1) = 0 \Rightarrow \sum_i s_i = \sum_i \mu$$

$$= N\mu \Rightarrow \mu = \frac{1}{N}\sum_i s_i$$

$$\therefore \boxed{\hat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N} s_i}$$

ML estimate of σ^2

- $\ln \mathcal{L} = -\frac{1}{2\underbrace{\sigma^2}_{a}}\sum_{i=1}^{N}(s_i-\mu)^2 - \frac{N}{2}\ln(2\pi\underbrace{\sigma^2}_{a})$

$\underbrace{\frac{\partial(\ln\mathcal{L})}{\partial\sigma}}_{\times}$

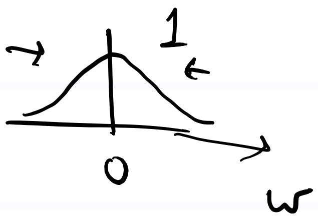
- let $\sigma^2 = a$

- $\hat{a} = \mathrm{argmax}_a - \frac{1}{2a}\sum_{i=1}^{N}(s_i-\mu)^2 - \frac{N}{2}\ln(2\pi a)$

$\frac{\partial(\ln\mathcal{L})}{\partial a} = +\frac{1}{2a^2}\sum_i(s_i-\mu)^2 - \frac{N}{2}\frac{1}{\cancel{2\pi}a}\cancel{2\pi} = 0 \quad \Rightarrow Na = \sum_i(s_i-\mu)^2$

$\therefore \boxed{\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(s_i-\hat{\mu}_{ML})^2}$

Estimating parameters with a prior belief



- For N independent measurements
  - $p(t_1, t_2, \ldots, t_N) = p(t_1)p(t_2)\ldots p(t_N)$

$p(w) = \mathcal{N}(w; 0, 1)$

- We want to estimate

$p(w \mid s_1, s_2, \ldots s_N) = ?$

  - $\boxed{\mu, \sigma = \underset{\mu,\sigma}{\operatorname{argmax}} \ln p(\mu, \sigma | t_1, \ldots, t_N)\big|_{t_1 = s_1, \ldots, t_N = s_N}}$

- This is called maximum a-posteriori estimate (MAP)

MAP for parameters of Gaussian pdf

- $p(\mu, \sigma | t_1, \dots, t_N) = \frac{p(t_1, \dots, t_N | \mu, \sigma) p(\mu, \sigma)}{p(t_1, \dots, t_N)}$
- Assume prior $p(\mu, \sigma)$
- $\ln p(\mu, \sigma | t_1, \dots, t_N) = \ln p(t_1, \dots, t_N | \mu, \sigma) + \ln p(\mu, \sigma)$

MAP Estimate of μ

$$\arg\max_{\mu} \; p(\mu|s_1,\dots s_N) = \left(\prod_i p(s_i|\mu)\right)\cdot p(\mu)$$

$$= \left(\prod_i \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{2}\, \underbrace{e^{-(s_i-\mu)^2/2\sigma^2}}_{1}\right)\cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma_0^2}}}_{4}\, \underbrace{e^{-(\mu-\mu_0)^2/2\sigma_0^2}}_{3}$$

$\frac{1}{Z}$

- Assume Gaussian prior

$$P(\mu) = \mathcal{N}(\mu;\mu_0,\sigma_0^2)$$

1 2 3 4

- $\ln \mathcal{L} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(s_i-\mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2 - \frac{1}{2}\ln(2\pi\sigma_0^2)$

$$\frac{\partial(\ln\mathcal{L})}{\partial\mu} = -\frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^{N}(s_i-\mu)\,\cancel{2}(-1) - \frac{1}{\cancel{2}\sigma_0^2}(\mu-\mu_0)\,\cancel{2} = 0 \Rightarrow \frac{\sum_i s_i}{\sigma^2} - \frac{\mu N}{\sigma^2} - \frac{\mu}{\sigma_0^2} + \frac{\mu_0}{\sigma_0^2} = 0$$

$$\Rightarrow \frac{\sum_i s_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} = \mu\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)$$

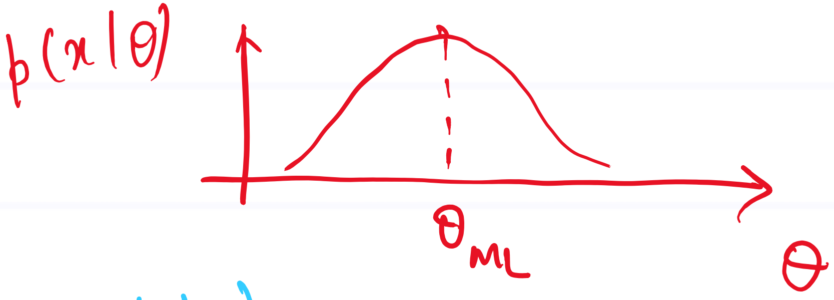
$$\therefore \boxed{\hat{\mu}_{MAP} = \left(\frac{\sum_{i=1}^{N} s_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \Big/ \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}$$

Summary

- I have samples
- consider r.v. x from where these samples are.
- Assume a model $p(x;\theta)$ with params θ
- Estimate model params θ

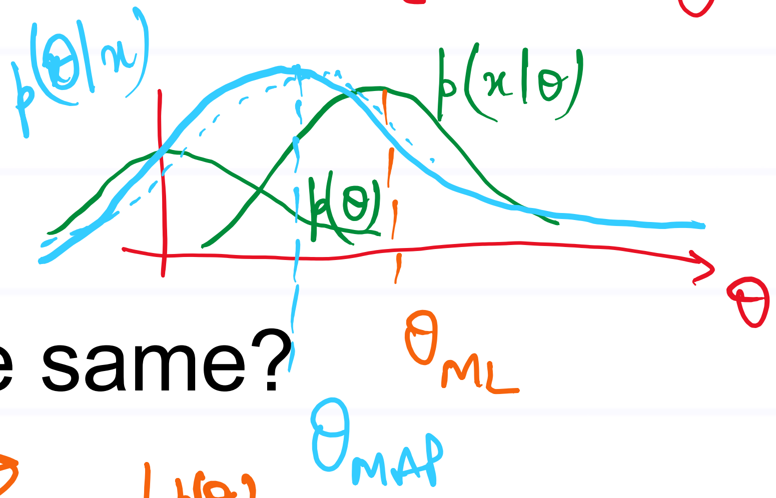
- $P(x;\theta)$
- $\hat{\theta}_{ML} = \text{argmax}_\theta P(x|\theta)$

↳ samples

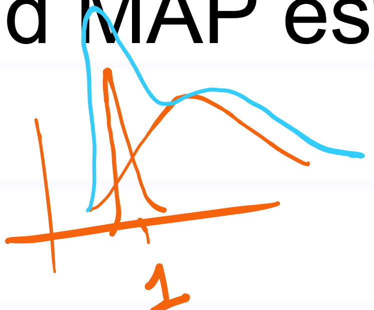


- $\hat{\theta}_{MAP} = \text{argmax}_\theta P(\theta|x)$ fixed

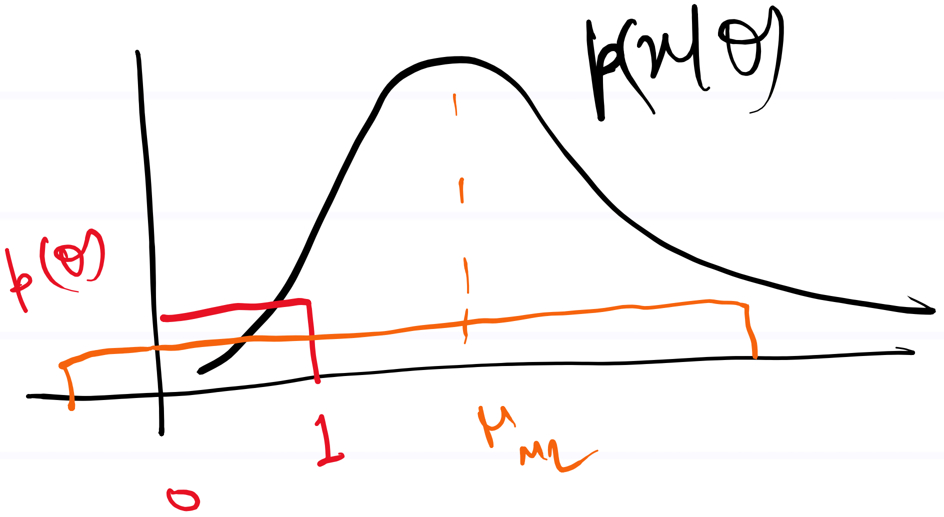
$$= \text{argmax}_\theta P(x|\theta)P(\theta)$$



When are ML and MAP estimate same?



When are ML and MAP estimates same?



$$p(\theta) = \mathcal{N}(\theta;\ 0,\ \sigma_\theta^2)$$

$$\sigma_\theta \gg 1$$

Reading

- PRML Section 1.2.5: Curve fitting re-visited
  - DIY

Shouldn't $\hat{\mu}$ be a random variable?

- MAP and ML estimate a single value of parameters
- Bayesian approach considers them as random variables and estimates their pdf

Bayesian estimation of μ

$$\mathcal{N}(\mu;\mu_0,\sigma_0^2)$$

- $p(\mu|t_1,\dots,t_N) = \dfrac{p(t_1,\dots,t_N|\mu)\,p(\mu)}{p(t_1,\dots,t_N)}$

$$\propto \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t_i-\mu)^2}{2\sigma^2}}\right) \times \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$$

$$= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^N} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\underbrace{-\sum_i (t_i-\mu)^2 \frac{1}{2\sigma^2} - (\mu-\mu_0)^2 \frac{1}{2\sigma_0^2}}_{\text{again quadratic in } \mu}}$$

again quadratic in μ

$\therefore$ $p(\mu|t_1 \dots t_N)$ is gaussian

Now find its μ' & σ'

Bayesian estimation of μ

Completing the squares to match with $-(\mu-\mu')^2/2\sigma'^2 = -\left(\frac{\mu^2}{2\sigma'^2} + \frac{\mu'^2}{2\sigma'^2} - \frac{2\mu\mu'}{2\sigma'^2}\right)$

$$-\sum_i (t_i-\mu)^2 \frac{1}{2\sigma^2} - (\mu-\mu_0)^2\frac{1}{2\sigma_0^2} = -\frac{1}{2\sigma^2}\left(\sum_i t_i^2 + N\mu^2 - 2\sum t_i\mu\right) - \frac{1}{2\sigma_0^2}\left(\mu^2+\mu_0^2-2\mu\mu_0\right)$$

collect μ^2 and μ terms

$$= \frac{-1}{2}\left[\mu^2\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}\right) + \mu\left(-\frac{2\sum t_i}{\sigma^2} - \frac{2\mu_0}{\sigma_0^2}\right) + \frac{\sum_i t_i^2}{\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2}\right]$$

$$\therefore \frac{1}{\sigma'^2} = \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}$$

$$\Rightarrow \sigma'^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}$$

$$+\left(\cancel{+}\frac{\cancel{2}\mu\mu'}{\cancel{2}\sigma'^2}\right) = +\frac{1}{\cancel{2}}\mu\left(\cancel{+}\frac{\cancel{2}\sum t_i}{\sigma^2} + \frac{\cancel{2}\mu_0}{\sigma_0^2}\right)$$

$$\frac{\mu\mu'}{\sigma'^2} = \mu\left(\frac{\sum t_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)$$

$$\therefore \mu' = \sigma'^2\cdot\left(\frac{\sum_i t_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) = \frac{\sum_i t_i\,\sigma_0^2 + \sigma^2\mu_0^2}{N\sigma_0^2+\sigma^2}$$

same as μ_{MAP}