pre install GPyTorch
L21

TAM 598 Lecture 22:

Gaussian Process Regression

Announcements:

- HW 6 covers lectures 21-23; due on Fri May 2 where we are:

* we have a Gaussian Process prior (an "infinite variate" gaussian) for unknown function $f(\cdot)$, given by $p(f(\cdot))$

* we are going to collect some data and use Bayes' rule to update our prior and obtain the posterior

$$p(f(\cdot) \mid D) \propto p(D \mid f(\cdot))\, p(f(\cdot))$$

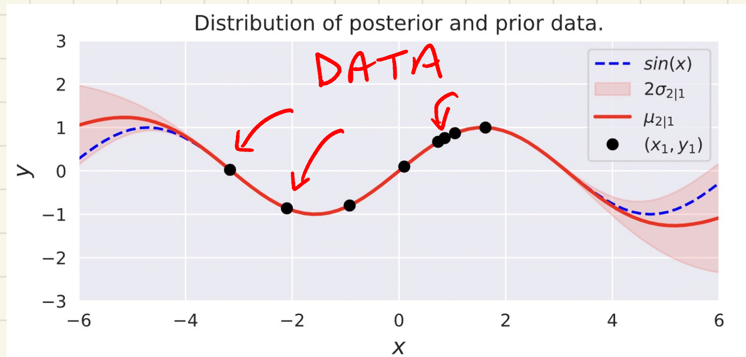
(we'll do this for an arbitrary n-variate Gaussian; the approach generalizes to the continuous infinite case)

$$p(f_1, f_2, \dots, f_n \mid D)$$

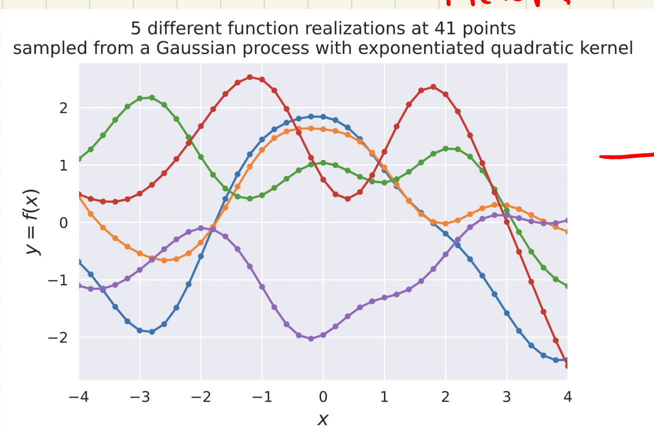
Gaussian Process ; a distribution over functions

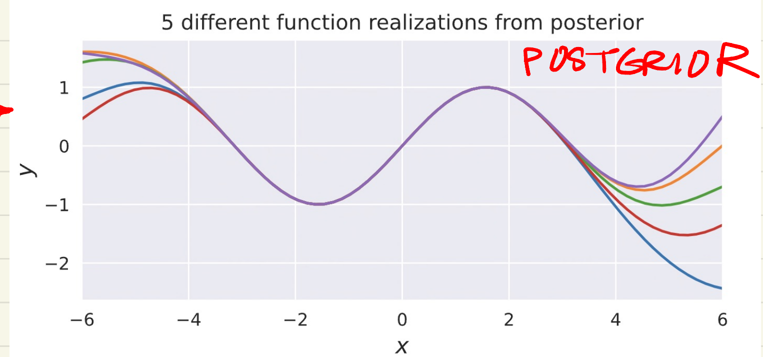
$$D = \{ (x_i, f_i), \ i = 1, \dots, N \}$$

$$\underbrace{p(f|D)}_{\text{posterior}} = \frac{p(D|f)\,\overbrace{p(f)}^{\text{prior}}}{p(D)}$$



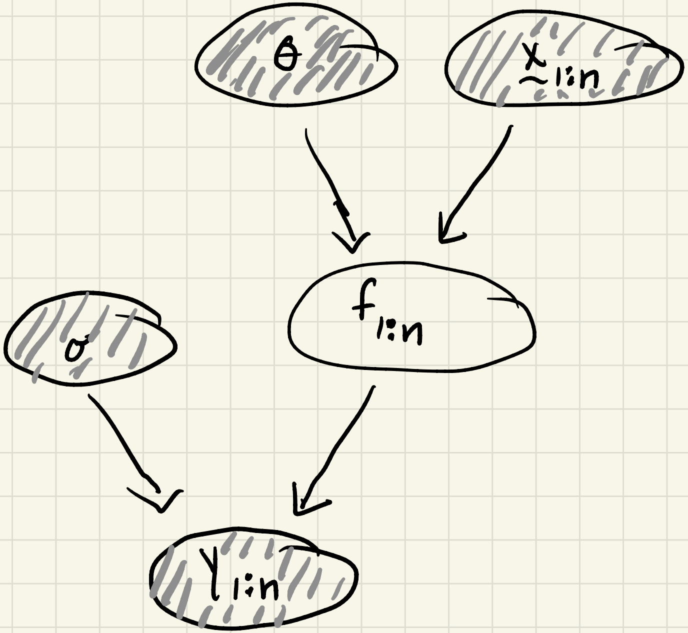
PRIOR





$f(\cdot) \sim GP(m(\cdot), k(\cdot, \cdot))$ so

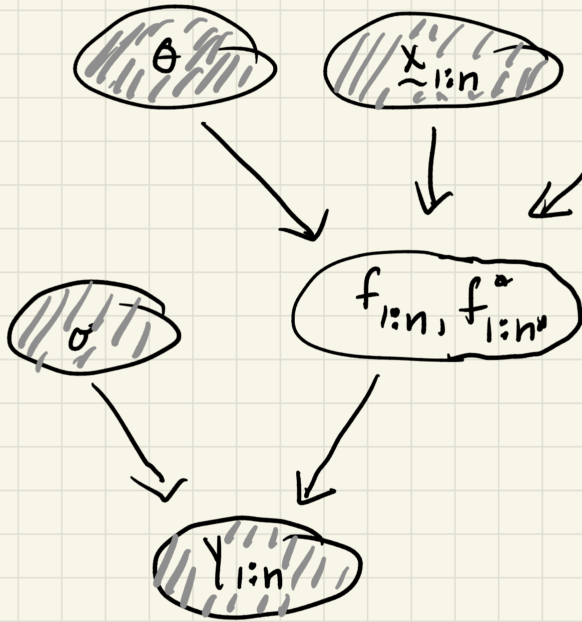
for any points $\underline{x}_{1:n} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ the joint pdf of $\underline{f}_{1:n} = (f(\underline{x}_1), \ldots, f(\underline{x}_n))$ is the multivariate gaussian



- fixed hyperparameters θ and σ

- arbitrary collection of test points, densely cover input space

and function values



$$p(\underline{f}^*_{1:n^*} \mid \underline{x}^*_{1:n^*}, D) = p(\underline{f}^*_{1:n^*} \mid \underline{x}^*_{1:n^*}, \underline{x}_{1:n}, \underline{y}_{1:n})$$

the result is a new Gaussian

$$p(\underline{f}^*_{1:n^*} \mid \underline{x}^*_{1:n^*}, D) = \mathcal{N}\left(\underline{f}^*_{1:n^*} \,\middle|\, \underline{m}_n(\underline{x}^*_{1:n^*}), \underline{K}_n\left(\underline{x}^*_{1:n^*}, \underline{x}^*_{1:n^*}\right)\right)$$

with posterior mean

and posterior covariance

point predictive distribution - to predict $f(\cdot)$ at a single point, have your test points $\underline{x}^*_{1:n^*}$ just be a single point $\underline{x}^*$. Then

but the predicted measurement outcome y^* at $\underline{x}^*$ is

HYPERPARAMETER OPTIMIZATION

- covariance function: θ
- measurement variance: σ^2

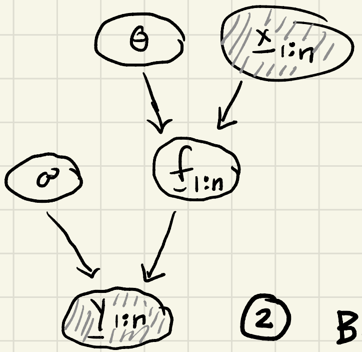
If we don't know them:

- start with priors $p(\theta)$, $p(\sigma)$
- Bayes Rule

$$p(\theta, \sigma \mid D) \propto p(D \mid \theta, \sigma)\, p(\theta)\, p(\sigma)$$

$$= \int p(\underline{y}_{1:n} \mid \underline{f}_{1:n}, \sigma)\, p(\underline{f}_{1:n} \mid \underline{x}_{1:n}, \theta)\, d\underline{f}_{1:n}\; p(\theta)\, p(\sigma)$$

- Maximum a posteriori estimation to evaluate



① joint pdf:

② Bayes Rule:

③ marginalize out the unobserved variable $\underline{f}_{1:n}$

Maximum a Posteriori (MAP) Estimate of Hyperparameters

$$p(\theta, \sigma | D) \approx \delta(\theta - \theta^*)\, \delta(\sigma - \sigma^*)$$

where θ^*, σ^* maximize $\log p(\theta, \sigma | D)$

so $\log p(\theta, \sigma | D) =$