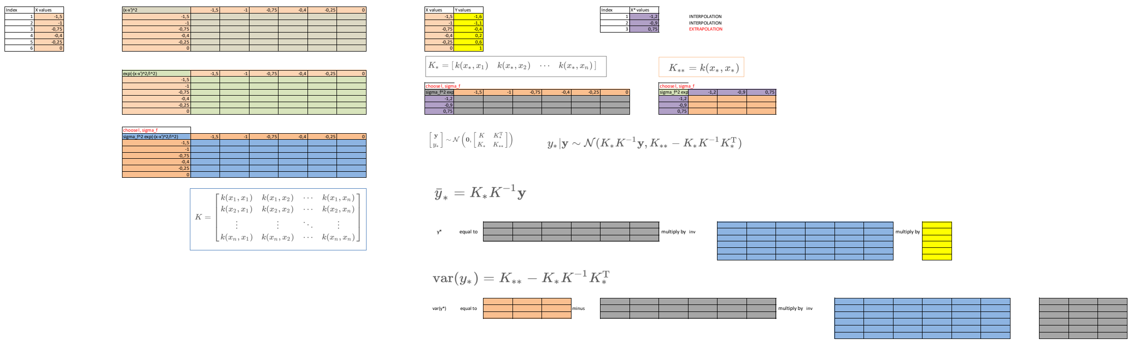
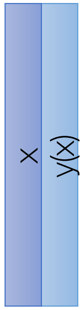
GP by hands

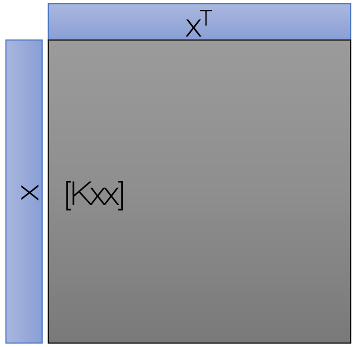
Prof. J. Morlier

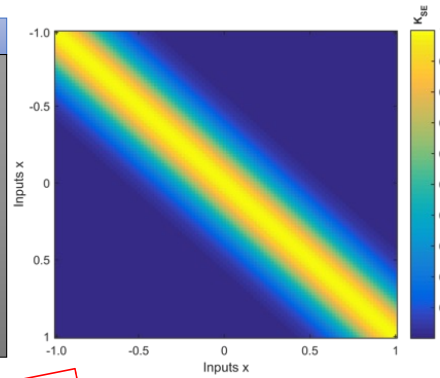


Matrix view of Gaussian Process



$$k(x,x') = \theta_1{}^2 \exp\left(-\frac{(x-x')^2}{2\theta_2{}^2}\right) =$$



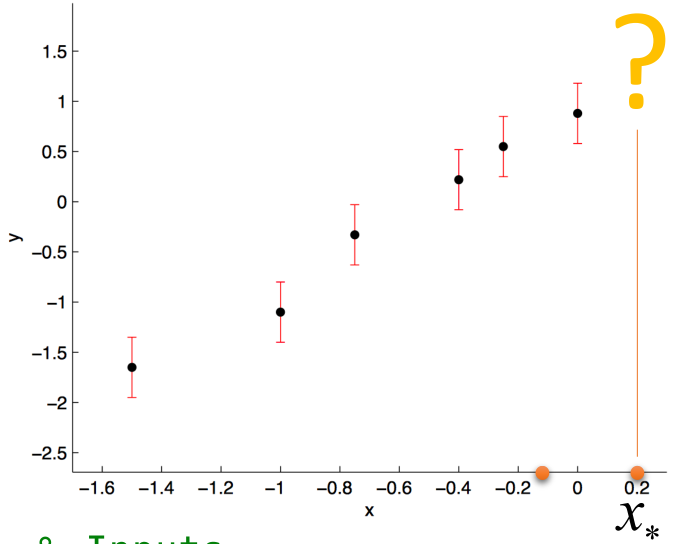


1/ Get your inputs/outputs data

2/ You wan to predict at x*

3/ Choose a Kernel/Construct Kxx and Hyperparameters tuning

Example




```
% Inputs
X = [-1.5 -1 -.75 -.4 -.25 0]';
% Outputs
y = [-1.6 -1.10 -.4 .2 0.6 1]';
```

(0.20, ?)

$$y \sim N(0, K)$$

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^2}{2l^2}\right] + \sigma_n \delta(x, x') \qquad \delta(x,x') = \begin{cases} 1, & x = x' \\ 0, & x \neq x' \end{cases}$$

$$K = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{bmatrix}$$

```
% 6*6
```

$$K_* = \begin{bmatrix} k(x_*,x_1) & k(x_*,x_2) & \cdots & k(x_*,x_n) \end{bmatrix} \qquad K_{**} = k(x_*, x_*)$$

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix}\right)$$

$$y_* | \mathbf{y} \sim \mathcal{N}(K_* K^{-1} \mathbf{y}, K_{**} - K_* K^{-1} K_*^{\mathrm{T}})$$

$$\bar{y}_* = K_* K^{-1} \mathbf{y} \qquad \mathrm{var}(y_*) = K_{**} - K_* K^{-1} K_*^{\mathrm{T}}$$

Gaussian Processes for Regression A Quick Introduction, M.Ebden, August 2008.

Step by step: l=1= θ_1, σ_f =sqrt(0.3)= θ_2

Inputs Table: The given X values.

Pairwise Differences: A matrix of $(x - x')^2$.

Exponential Term: Computing $\exp(-(x - x')^2/l^2)$.

Final Kernel Matrix: Multiplying by σ_f^2.

$$k(x, x') = \theta_1{}^2 \exp\left(-\frac{(x - x')^2}{2\theta_2{}^2}\right)$$

$$\theta = [\theta_1;\ \theta_2]$$

Step by step:

Inputs Table: The given X values.

Pairwise Differences: A matrix of $(x - x')^2$.

Exponential Term: Computing $\exp(-(x - x')^2 / l^2)$.

Final Kernel Matrix: Multiplying by σ_f^2.

Index	X values
1	-1,5
2	-1
3	-0,75
4	-0,4
5	-0,25
6	0

Step by step:

Inputs Table: The given X values.

Pairwise Differences: A matrix of $(x - x')^2$.

Exponential Term: Computing $\exp(-(x - x')^2 / l^2)$.

Final Kernel Matrix: Multiplying by σ_f^2.

```
X = np.array([-1.5, -1, -0.75, -0.4, -0.25, 0])
l = 1  # Length scale
sig_f = 3**(1/4)  # Since sigma_f^2 = sqrt(3)
```

(x-x')^2	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5	0	0,25	0,5625	1,21	1,5625	2,25
-1	0,25	0	0,0625	0,36	0,5625	1
-0,75	0,5625	0,0625	0	0,1225	0,25	0,5625
-0,4	1,21	0,36	0,1225	0	0,0225	0,16
-0,25	1,5625	0,5625	0,25	0,0225	0	0,0625
0	2,25	1	0,5625	0,16	0,0625	0

Step by step:

Inputs Table: The given X values.

Pairwise Differences: A matrix of $(x - x')^2$.

Exponential Term: Computing $\exp(-(x - x')^2/l^2)$.

Final Kernel Matrix: Multiplying by σ_f^2.

```
pairwise_diffs = np.zeros((len(X), len(X)))
exp_terms = np.zeros((len(X), len(X)))

for i in range(len(X)):
   for j in range(len(X)):
      pairwise_diffs[i, j] = (X[i] - X[j]) ** 2
      exp_terms[i, j] = np.exp(-pairwise_diffs[i, j] / (l ** 2))

final_kernel = (sig_f ** 2) * exp_terms
```

exp(-(x-x')^2/l^2)	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5	1	0,778801	0,569783	0,298197	0,209611	0,105399
-1	0,778801	1	0,939413	0,697676	0,569783	0,367879
-0,75	0,569783	0,939413	1	0,884706	0,778801	0,569783
-0,4	0,298197	0,697676	0,884706	1	0,977751	0,852144
-0,25	0,209611	0,569783	0,778801	0,977751	1	0,939413
0	0,105399	0,367879	0,569783	0,852144	0,939413	1

Step by step:

Inputs Table: The given X values.

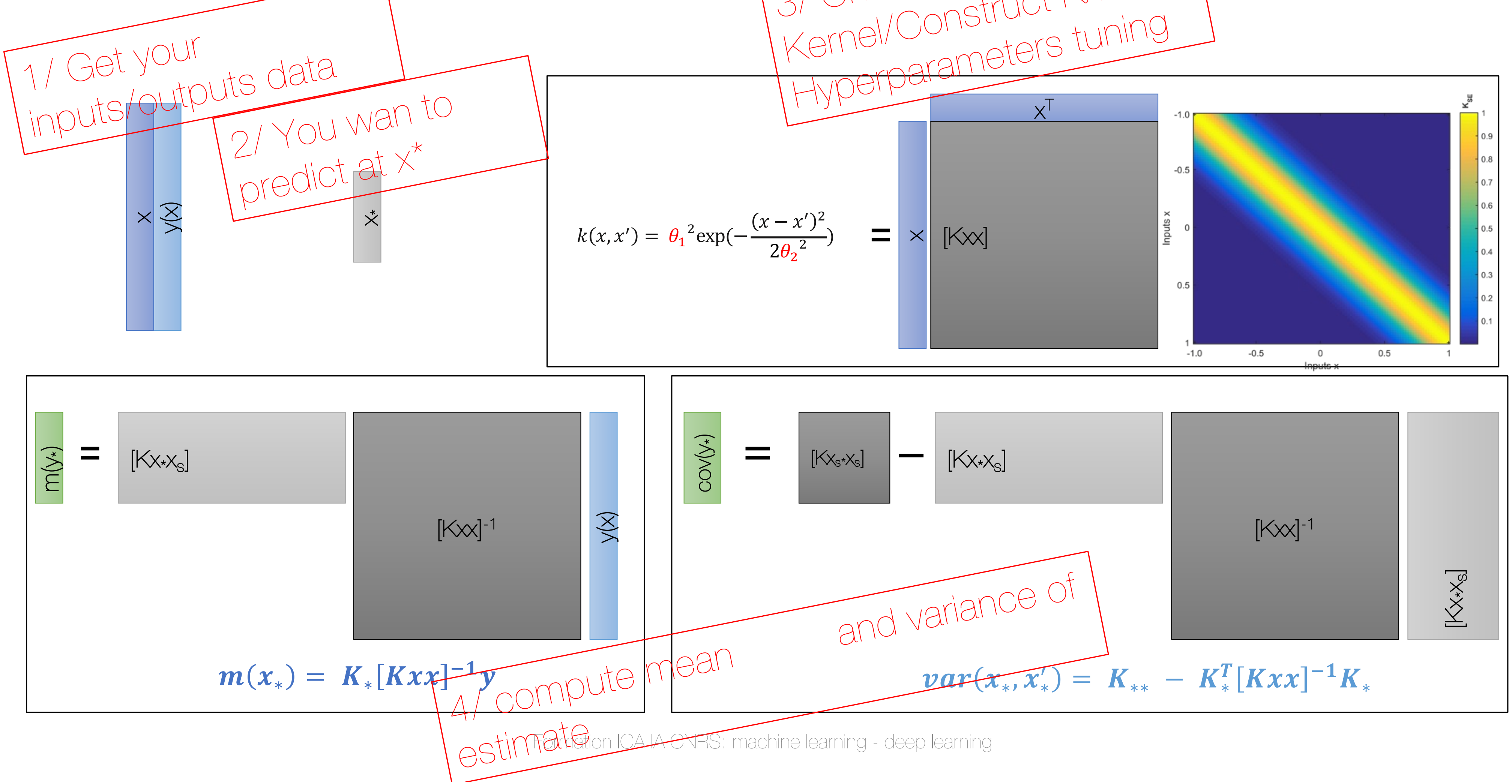
Pairwise Differences: A matrix of $(x - x')^2$.

Exponential Term: Computing $\exp(-(x - x')^2 / l^2)$.

Final Kernel Matrix: Multiplying by σ_f^2.

sigma_f^2 exp(-(x-x')^2/l^2)	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5	1,732051	1,348923	0,986893	0,516493	0,363058	0,182557
-1	1,348923	1,732051	1,627111	1,208411	0,986893	0,637186
-0,75	0,986893	1,627111	1,732051	1,532356	1,348923	0,986893
-0,4	0,516493	1,208411	1,532356	1,732051	1,693515	1,475956
-0,25	0,363058	0,986893	1,348923	1,693515	1,732051	1,627111
0	0,182557	0,637186	0,986893	1,475956	1,627111	1,732051

Matrix view of Gaussian Process

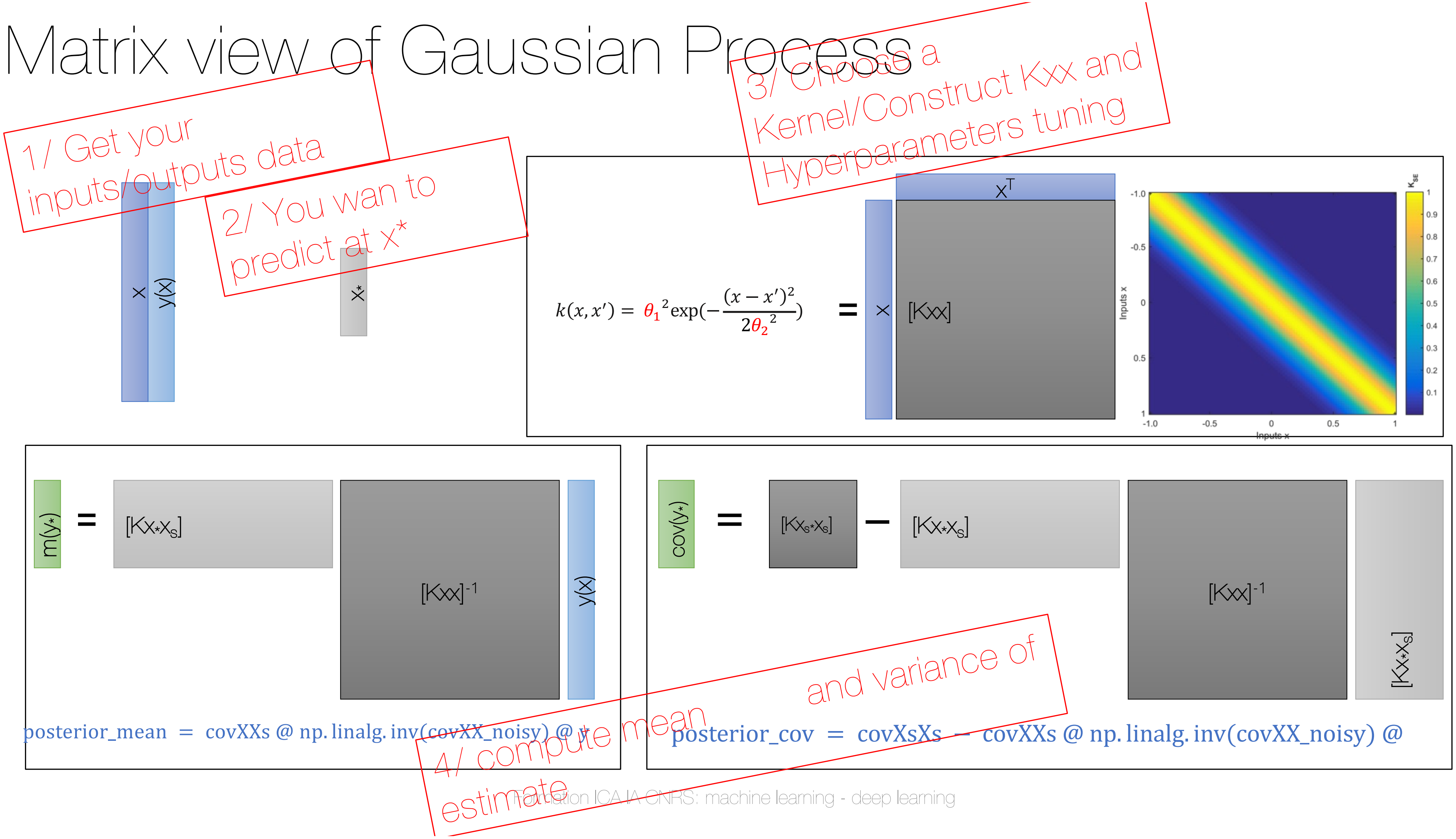


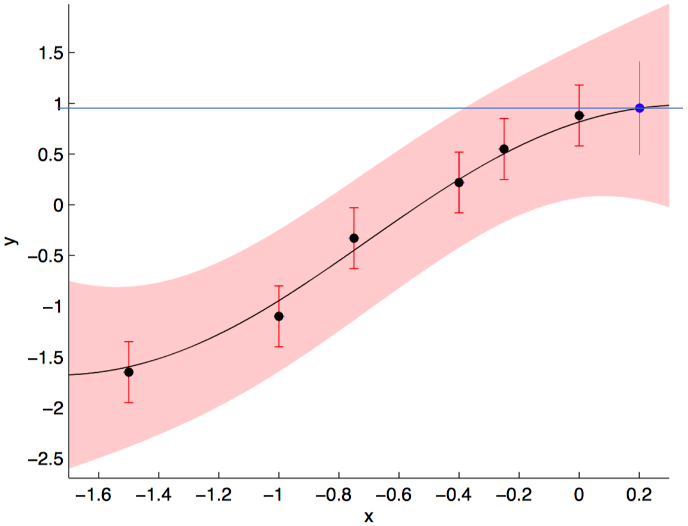
$$k(x,x') = \theta_1{}^2 \exp\left(-\frac{(x-x')^2}{2\theta_2{}^2}\right)$$

$$m(x_*) = K_*[Kxx]^{-1}y$$

$$var(x_*, x'_*) = K_{**} - K_*^T[Kxx]^{-1}K_*$$

Matrix view of Gaussian Process





$$\sigma_n = 0.3$$

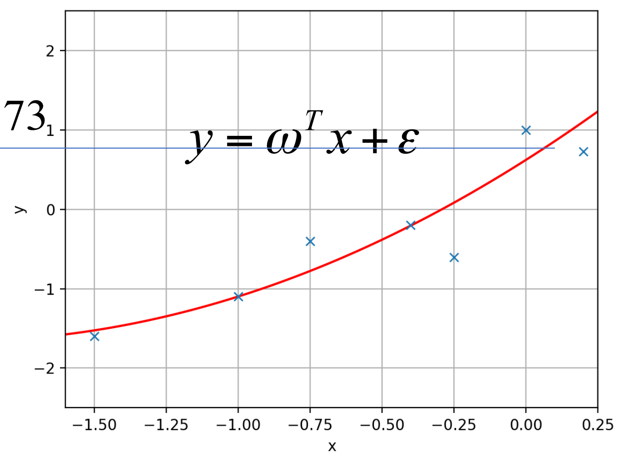
$$K = \begin{bmatrix} 1.70 & 1.42 & 1.21 & 0.87 & 0.72 & 0.51 \\ 1.42 & 1.70 & 1.56 & 1.34 & 1.21 & 0.97 \\ 1.21 & 1.56 & 1.70 & 1.51 & 1.42 & 1.21 \\ 0.87 & 1.34 & 1.51 & 1.70 & 1.59 & 1.48 \\ 0.72 & 1.21 & 1.42 & 1.59 & 1.70 & 1.56 \\ 0.51 & 0.97 & 1.21 & 1.48 & 1.56 & 1.70 \end{bmatrix}$$

$$K_{**} = 1.70 \qquad K_* = \begin{bmatrix} 0.38 & 0.79 & 1.03 & 1.35 & 1.46 & 1.58 \end{bmatrix}$$

$$\bar{y}_* = 0.95 \qquad \text{var}(y_*) = 0.21$$

$$\hat{y}_* = 0.73$$

$$y = \omega^T x + \varepsilon$$



http://htmlpreview.github.io/?https://github.com/jomorlier/mdocourse/blob/master/GP_Tutorial/GP_Tutorial.html

Index	X values
1	-1,5
2	-1
3	-0,75
4	-0,4
5	-0,25
6	0

(x-x')^2	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5						
-1						
-0,75						
-0,4						
-0,25						
0						

exp(-(x-x')^2/l^2)	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5						
-1						
-0,75						
-0,4						
-0,25						
0						

choose l, sigma_f

sigma_f^2 exp(-(x-x')^2/l^2)	-1,5	-1	-0,75	-0,4	-0,25	0
-1,5						
-1						
-0,75						
-0,4						
-0,25						
0						

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

X values	Y values
-1,5	-1,6
-1	-1,1
-0,75	-0,4
-0,4	0,2
-0,25	0,6
0	1

Index	X* values	
1	-1,2	INTERPOLATION
2	-0,9	INTERPOLATION
3	0,75	EXTRAPOLATION

$$K_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \cdots & k(x_*, x_n) \end{bmatrix}$$

$$K_{**} = k(x_*, x_*)$$

choose l, sigma_f

sigma_f^2 exp	-1,5	-1	-0,75	-0,4	-0,25	0
-1,2						
-0,9						
0,75						

choose l, sigma_f

sigma_f^2 exp	-1,2	-0,9	0,75
-1,2			
-0,9			
0,75			

Problem of inv(K) when K is large, or too close point → ill conditionned

Solution? Partitioned Inverse

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix}\right)$$

$$y_* | \mathbf{y} \sim \mathcal{N}(K_* K^{-1} \mathbf{y}, K_{**} - K_* K^{-1} K_*^{\mathrm{T}})$$

$$\bar{y}_* = K_* K^{-1} \mathbf{y}$$

y* equal to multiply by inv multiply by

$$\mathrm{var}(y_*) = K_{**} - K_* K^{-1} K_*^{\mathrm{T}}$$

var(y*) equal to minus multiply by inv

Optimizing Marginal Likelihood (ML)

$$\text{ML} = log\big(p(y|X,\theta)\big) = -\frac{1}{2}y^T K^{-1} y - \frac{1}{2}log|K| - \frac{n}{2}log(2\pi)$$

- It is a combination of **data-fit term**, a **complexity penalty** term and a **normalization term**

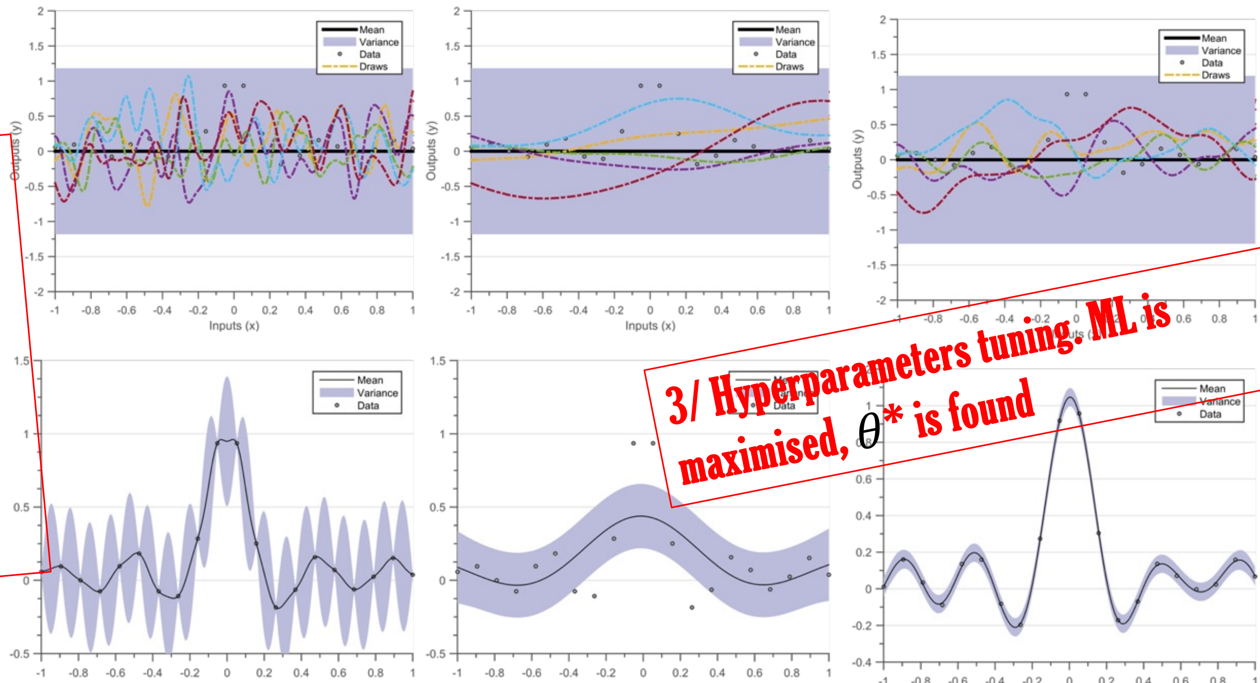
$$k(x,x') = \theta_1{}^2 \exp\left(-\frac{(x-x')^2}{2\theta_2{}^2}\right)$$

Only two hyperparameters:

→ **The lengthscale θ_2 or ℓ determines the length of the 'wiggles' in your function.**

→ **The output variance $\theta_1{}^2$ or σ^2 determines the average distance of your function away from its mean. It's just a scale factor.**

→ **A third hyperparameter θ_3 or σ_n^2 is often used (noise) $GP(0, K+\sigma_n^2 I)$**



ML = -8.2

ML = -35.3

ML = 6,04

3/ Hyperparameters tuning: ML is maximised, θ^* is found

DOCC_D3_C2