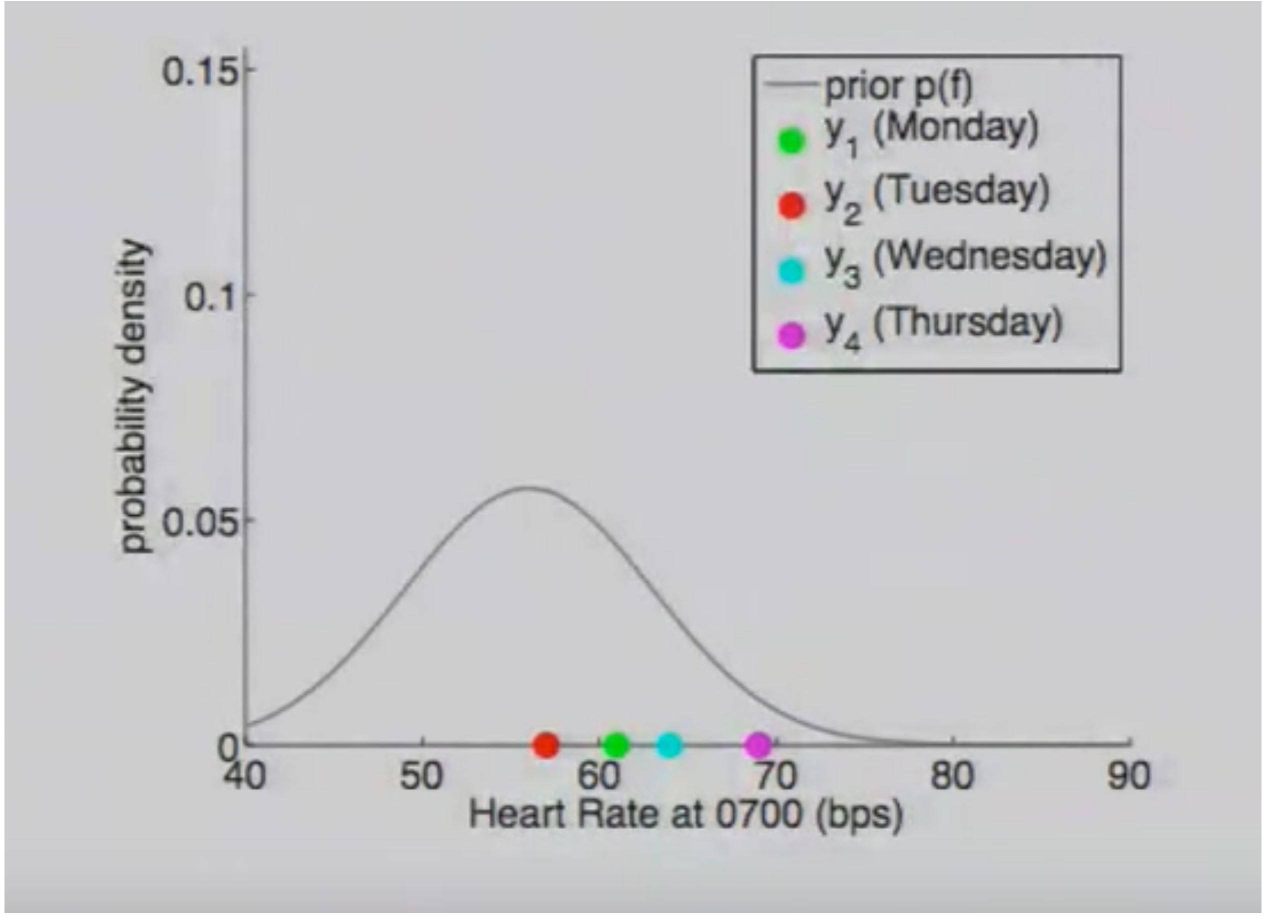
Introduction to Gaussian Processes

Sources:

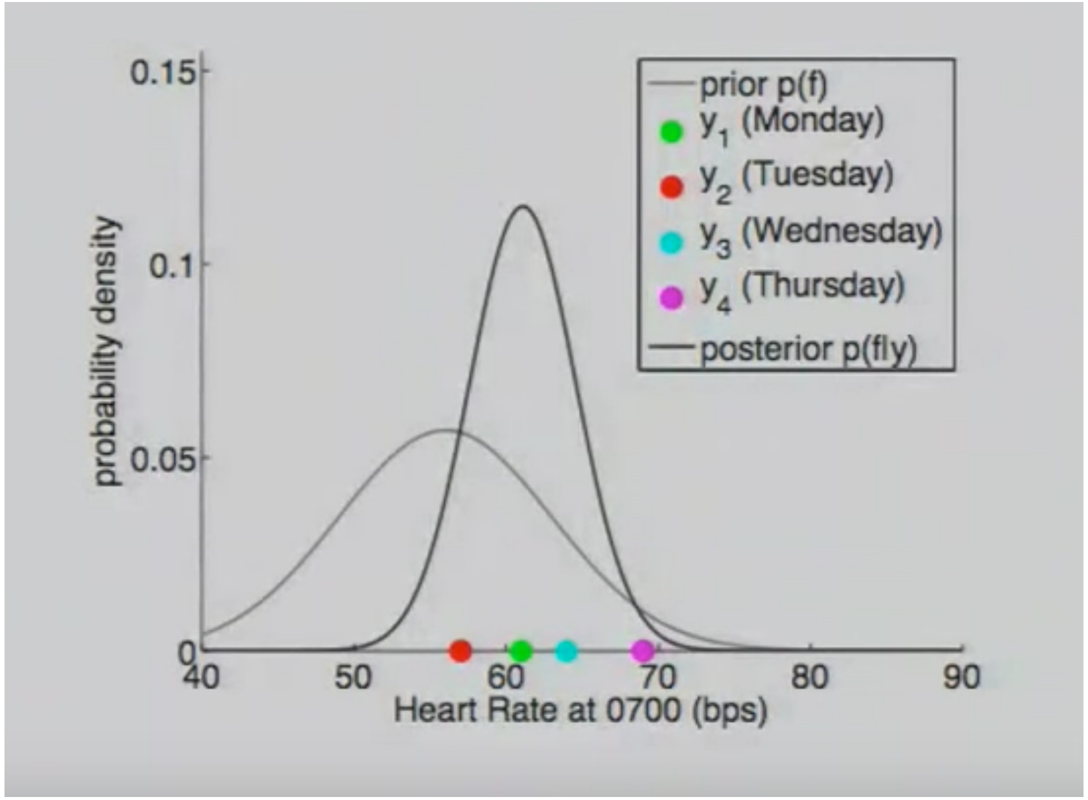
Rasmussen & Williams http://www.gaussianprocess.org/gpml/chapters/

John Cunningham: https://www.youtube.com/watch?v=BS4Wd5rwNwE&t=2348s
(this lecture)

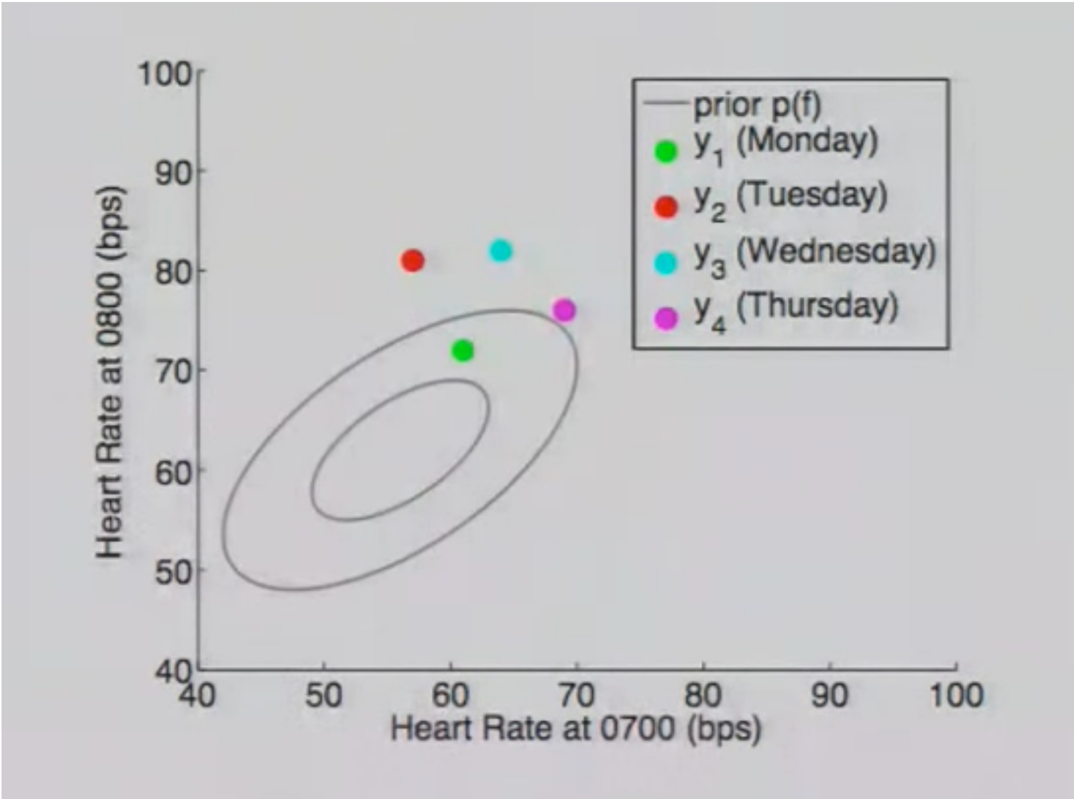
Measuring one’s heart rate at 7am on successive days

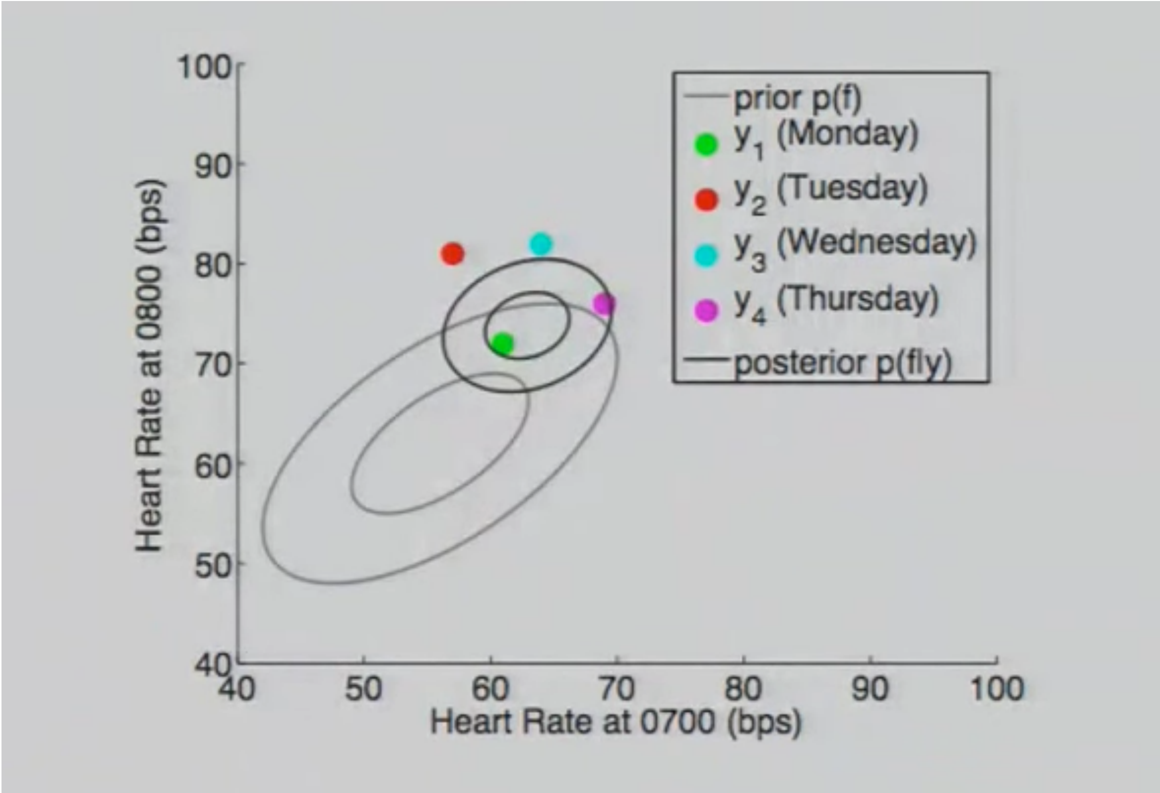


Use Bayesian reasoning to compute a posteriori distribution



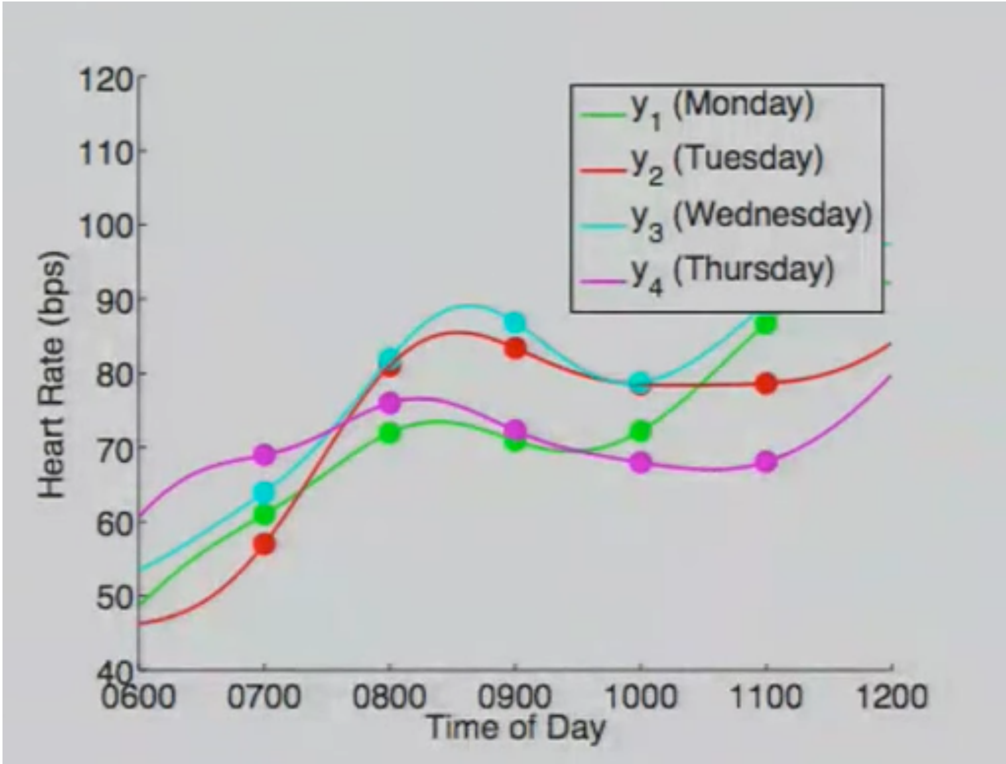
Extend analysis to two times of day

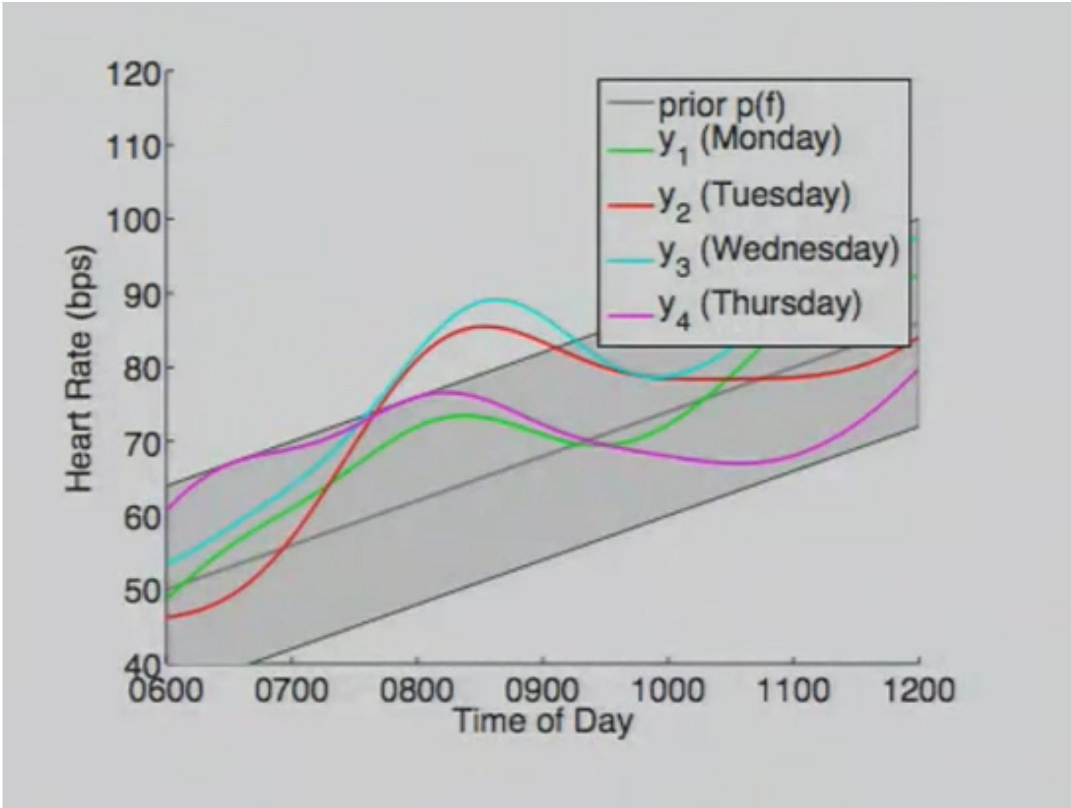




prior p(f)
y1 (Monday)
y2 (Tuesday)
y3 (Wednesday)
y4 (Thursday)
posterior p(f|y)
Heart Rate at 0800 (bps)
Heart Rate at 0700 (bps)
40
50
60
70
80
90
100

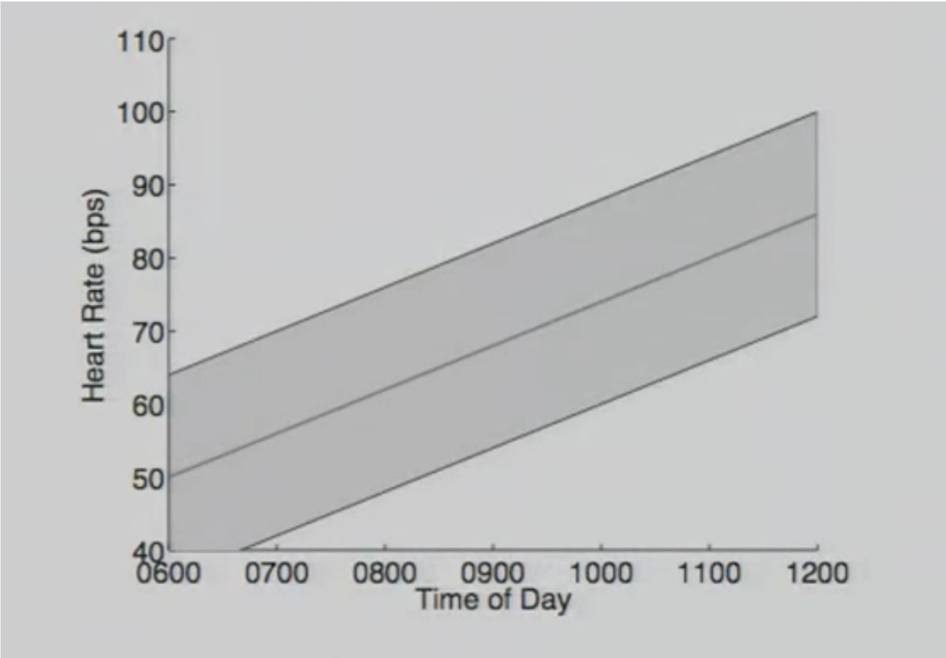
Plot the data as a function of time

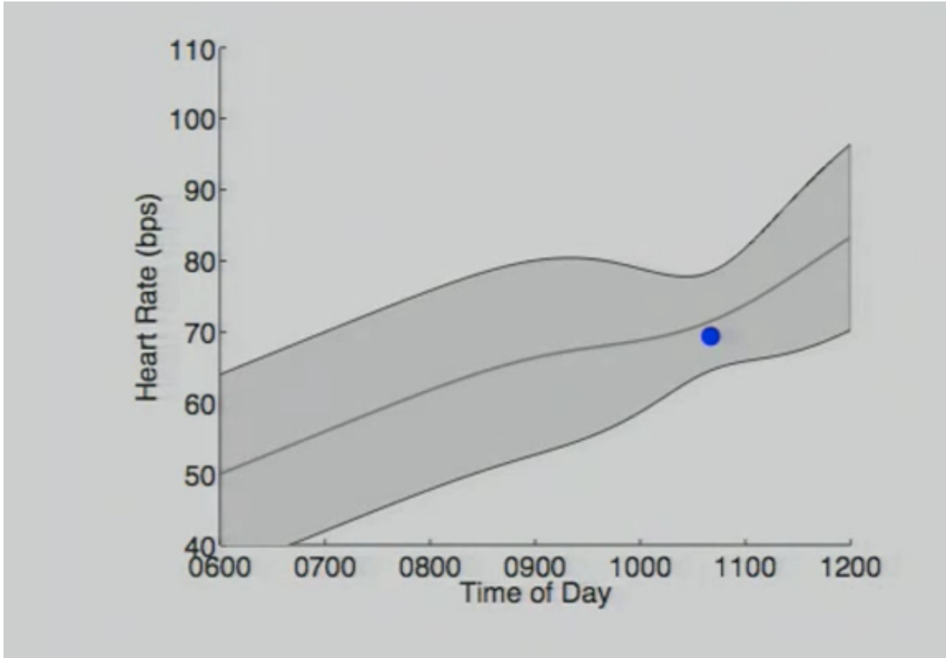




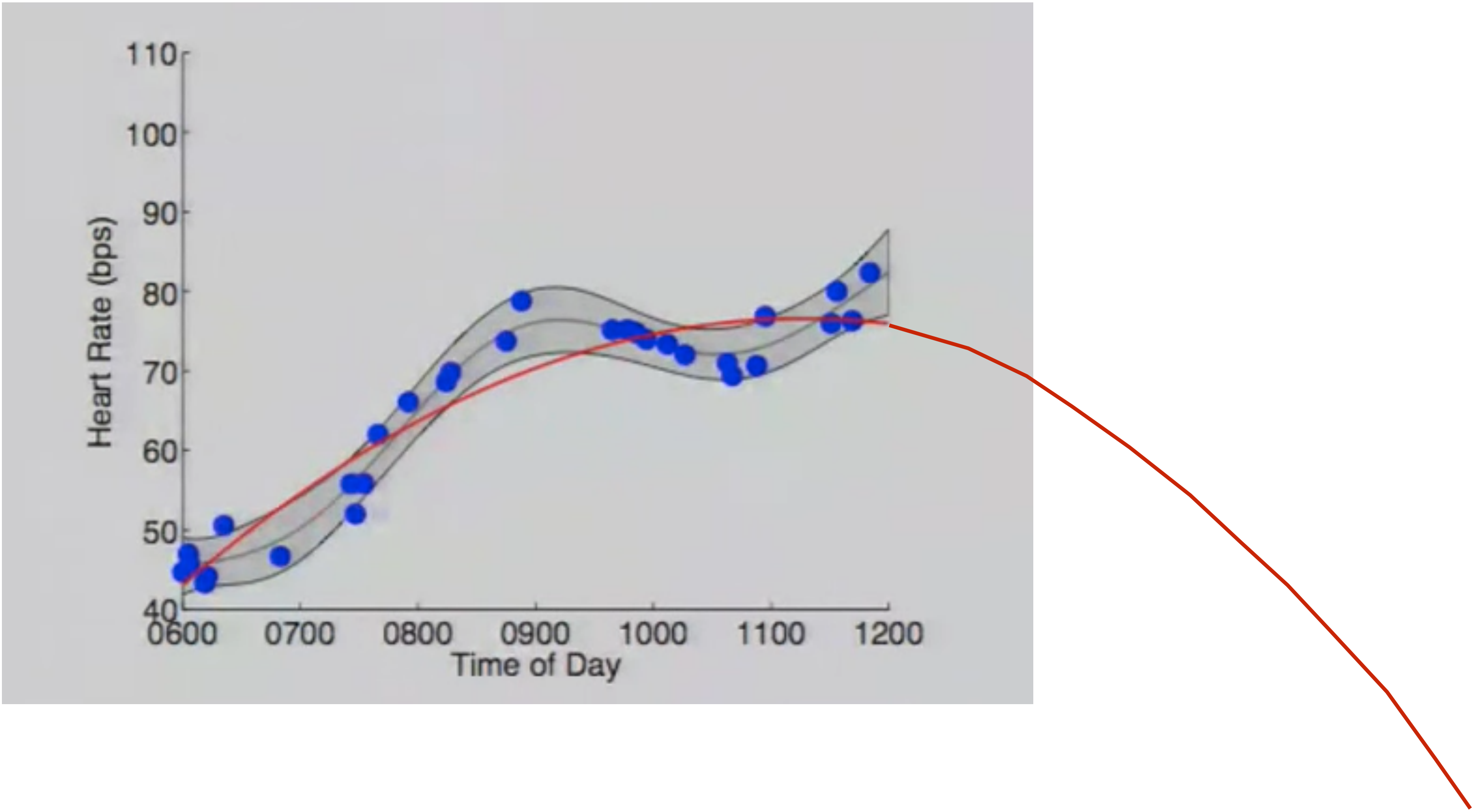
Heart Rate (bps)
120
110
100
90
80
70
60
50
40
0600
0700
0800
0900
1000
1100
1200
Time of Day
prior p(f)
y1 (Monday)
y2 (Tuesday)
y3 (Wednesday)
y4 (Thursday)

Data measurement refines the a posteriori distribution





An argument against parametric models e.g parabola



Gaussian definition

$f \in \mathbb{R}^n$ is normally distributed if
$p(f) = (2\pi)^{-\frac{n}{2}} |K|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(f-m)^T K^{-1}(f-m)\right\}$

for mean vector $m \in \mathbb{R}^n$ and positive semidefinite covariance matrix $K \in \mathbb{R}^{n \times n}$

shorthand: $f \sim \mathcal{N}(m, K)$

Gaussian **Process** definition

f is a Gaussian process if $f(t) = [f(t_1), ..., f(t_n)]'$ has a multivariate normal distribution for all $t = [t_1, ..., t_n]'$:

$$f(t) \sim \mathcal{N}(m(t), K(t, t))$$

K must be positive definite

$$v^T K(t, t) v = \sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij} v_i v_j = \sum_{i=1}^{n} \sum_{j=1}^{n} k(t_i, t_j) v_i v_j \geq 0$$

Can use a **kernel function** as a prior for the GP

$$k(t_i, t_j) = \sigma_f^2 \exp\left\{-\frac{1}{2\ell^2}(t_i - t_j)^2\right\}$$

From kernel to covariance matrix

- Choose some *hyperparameters*: $\sigma_f = 7$, $\ell = 100$

$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 49.0 & 29.7 & 00.2 \\ 29.7 & 49.0 & 03.6 \\ 00.2 & 03.6 & 49.0 \end{bmatrix}$$

- Choose some *hyperparameters*: $\sigma_f = 7$, $\ell = 500$

$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 49.0 & 48.0 & 39.5 \\ 48.0 & 49.0 & 44.1 \\ 39.5 & 44.1 & 49.0 \end{bmatrix}$$

- Choose some *hyperparameters*: $\sigma_f = 7$, $\ell = 50$

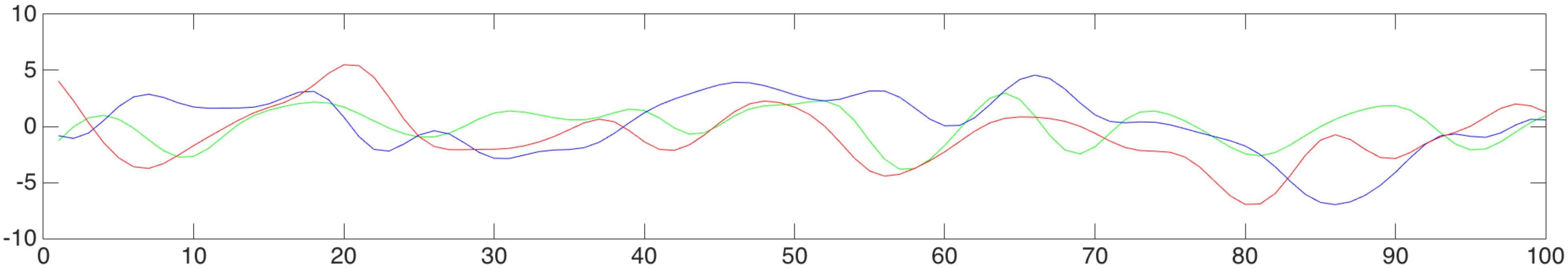
$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 49.0 & 06.6 & 00.0 \\ 06.6 & 49.0 & 00.0 \\ 00.0 & 00.0 & 49.0 \end{bmatrix}$$

If f and y are jointly Gaussian:

$$\begin{bmatrix} f \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_f \\ m_y \end{bmatrix}, \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix} \right)$$

Then:

$$f|y \sim \mathcal{N}\left(K_{fy}K_{yy}^{-1}(y - m_y) + m_f \;,\; K_{ff} - K_{fy}K_{yy}^{-1}K_{fy}^T\right)$$




```
function [ Km ] = GaussianF( numPts,s,l )
%Priors for GP

Km=zeros(numPts,numPts);
for i=1:numPts
    for j=1:numPts
        Km(i,j)=(s^2)*exp(-(0.5/l^2)*(i-j)^2);
    end
end
end


 F=chol(Km)'*randn(100,1);
```

prior (or latent) $f \sim \mathcal{N}(m_f, K_{ff})$

additive iid noise $n \sim \mathcal{N}(0, \sigma_n^2 I)$

let $y = f + n$, then:

$$p(y, f) = p(y|f)p(f) = \mathcal{N}\left(\begin{bmatrix} f \\ y \end{bmatrix}; \begin{bmatrix} m_f \\ m_y \end{bmatrix}, \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix} \right)$$

where (in this case):

$$K_{fy} = E[(f - m_f)(y - m_y)^T] = K_{ff} \qquad K_{yy} = K_{ff} + \sigma_n^2 I$$

prior (or latent) $f \sim \mathcal{GP}(m_f, k_{ff})$

additive iid noise $n \sim \mathcal{GP}(0, \sigma_n^2 \delta)$

let $y = f + n$, then:

$$p(y(t), f(t)) = p(y|f)p(f) = \mathcal{N}\left(\begin{bmatrix} f \\ y \end{bmatrix}; \begin{bmatrix} m_f \\ m_y \end{bmatrix}, \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix}\right)$$

Conditioning on data gave us:

$$f|y \sim \mathcal{N}\left(K_{fy}K_{yy}^{-1}(y - m_y) + m_f \ , \ K_{ff} - K_{fy}K_{yy}^{-1}K_{fy}^T\right)$$

then $E[f|y] = K_{fy}K_{yy}^{-1}(y - m_y) + m_f$ (MAP, posterior mean, ...)

Predict data observations y^*:

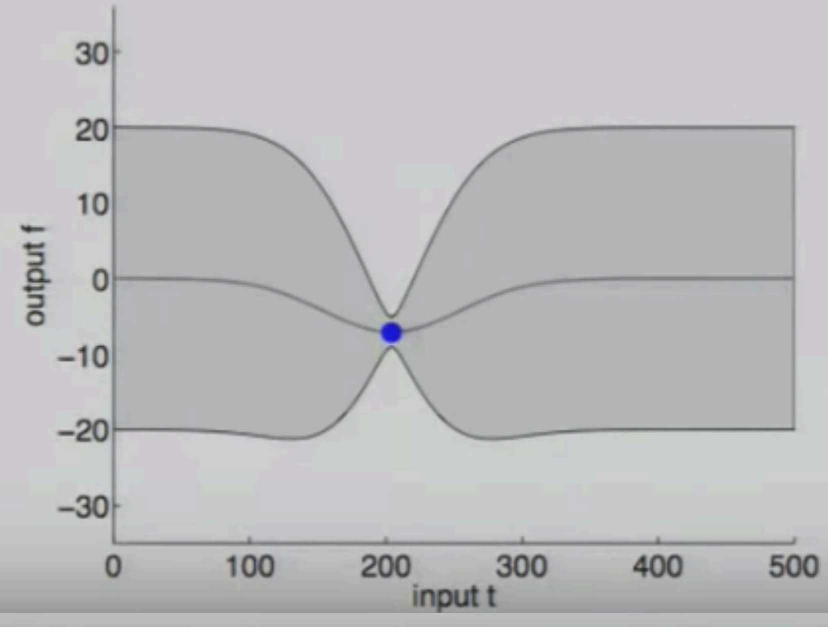
$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_y \\ m_{y^*} \end{bmatrix}, \begin{bmatrix} K_{yy} & K_{y^*y} \\ K_{y^*y}^T & K_{y^*y^*} \end{bmatrix}\right)$$

no different:

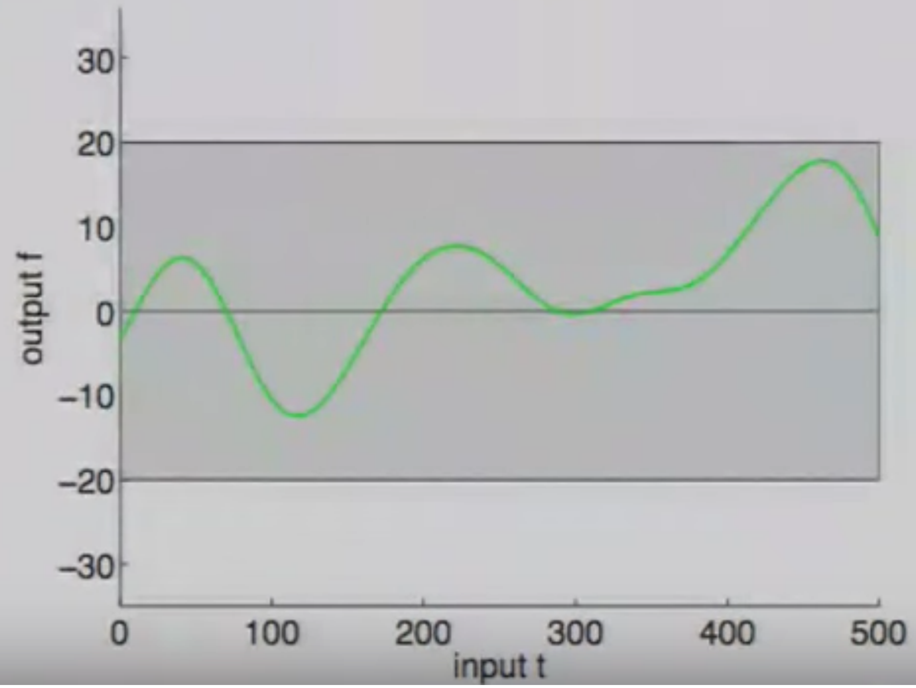
$$y^*|y \sim \mathcal{N}\left(K_{y^*y}K_{yy}^{-1}(y - m_y) + m_{y^*} \ , \ K_{y^*y^*} - K_{y^*y}K_{yy}^{-1}K_{y^*y}^T\right)$$

- Use conditioning to update the posterior:

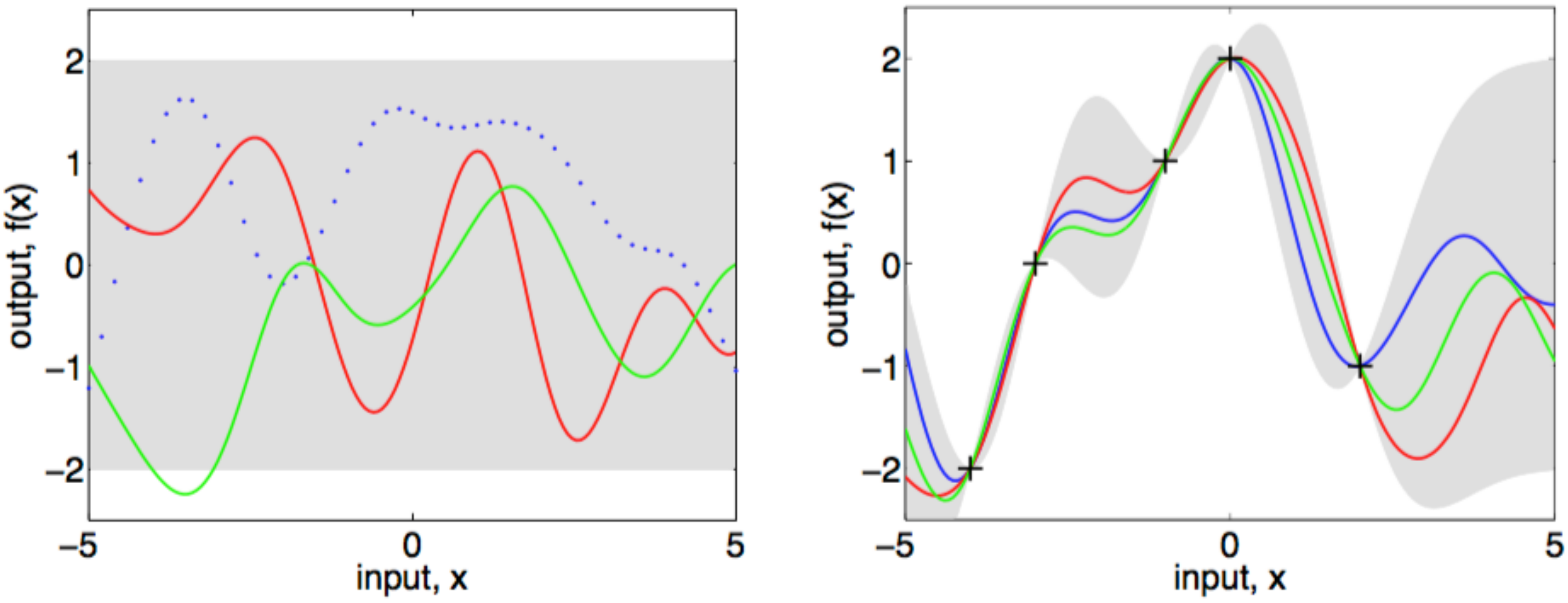
$$f|y(204) \sim \mathcal{N}\left(K_{fy}K_{yy}^{-1}(y(204) - m_y)\ ,\ K_{ff} - K_{fy}K_{yy}^{-1}K_{fy}^{T}\right)$$



- $f \sim \mathcal{GP}(0, k_{ff})$, where $k_{ff}(t_i, t_j) = \sigma_f^2 \exp\left\{-\frac{1}{2\ell^2}(t_i - t_j)^2\right\}$
- $y|f \sim \mathcal{GP}(f, k_{nn})$, where $k_{nn}(t_i, t_j) = \sigma_n^2 \delta(t_i - t_j)$
- $y \sim \mathcal{GP}(0, k_{yy})$, where $k_{yy}(t_i, t_j) = k_{ff}(t_i, t_j) + k_{nn}(t_i, t_j)$
- We choose $\sigma_f = 10$, $\ell = 50$, $\sigma_n = 1$
- A draw from f:

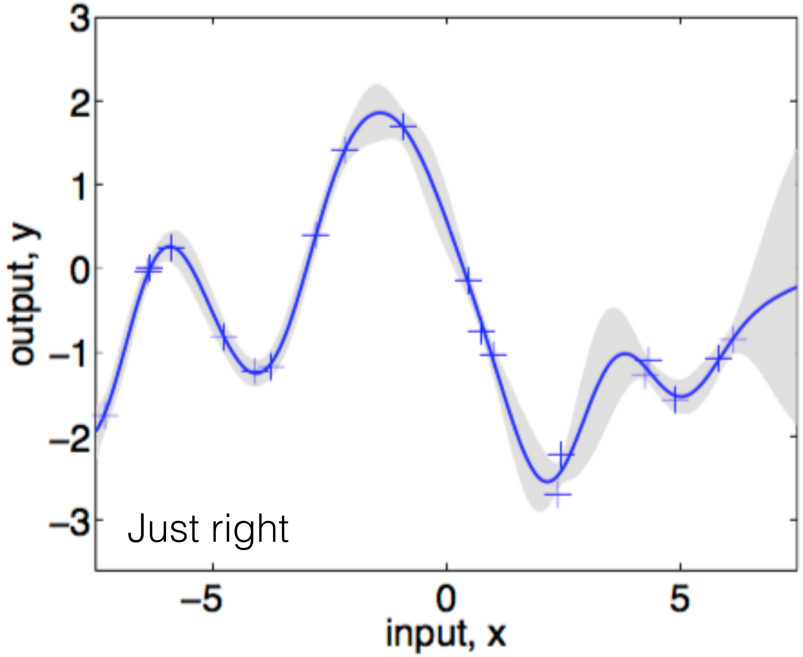


An impractical but intuitively appealing way to think about
Bayesian reasoning in the GP in the noiseless case:
Of all the possible functions, just keep the ones that fit the data

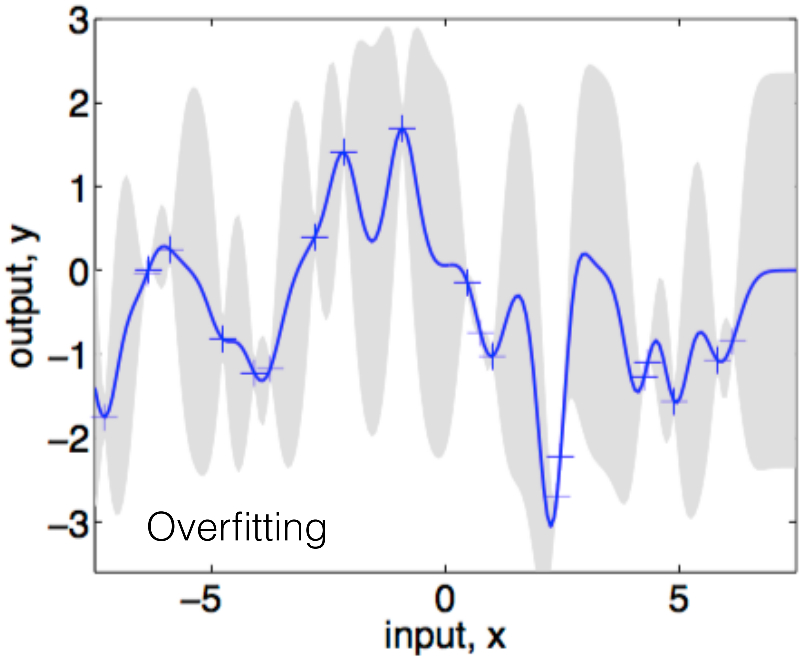


(a), prior (b), posterior

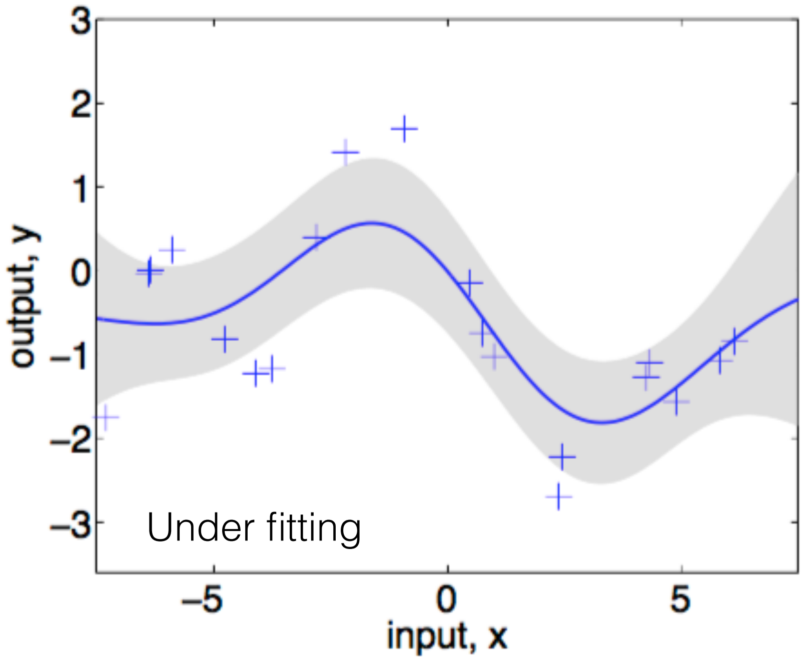
A practical problem: choosing the best set of hyperparameters



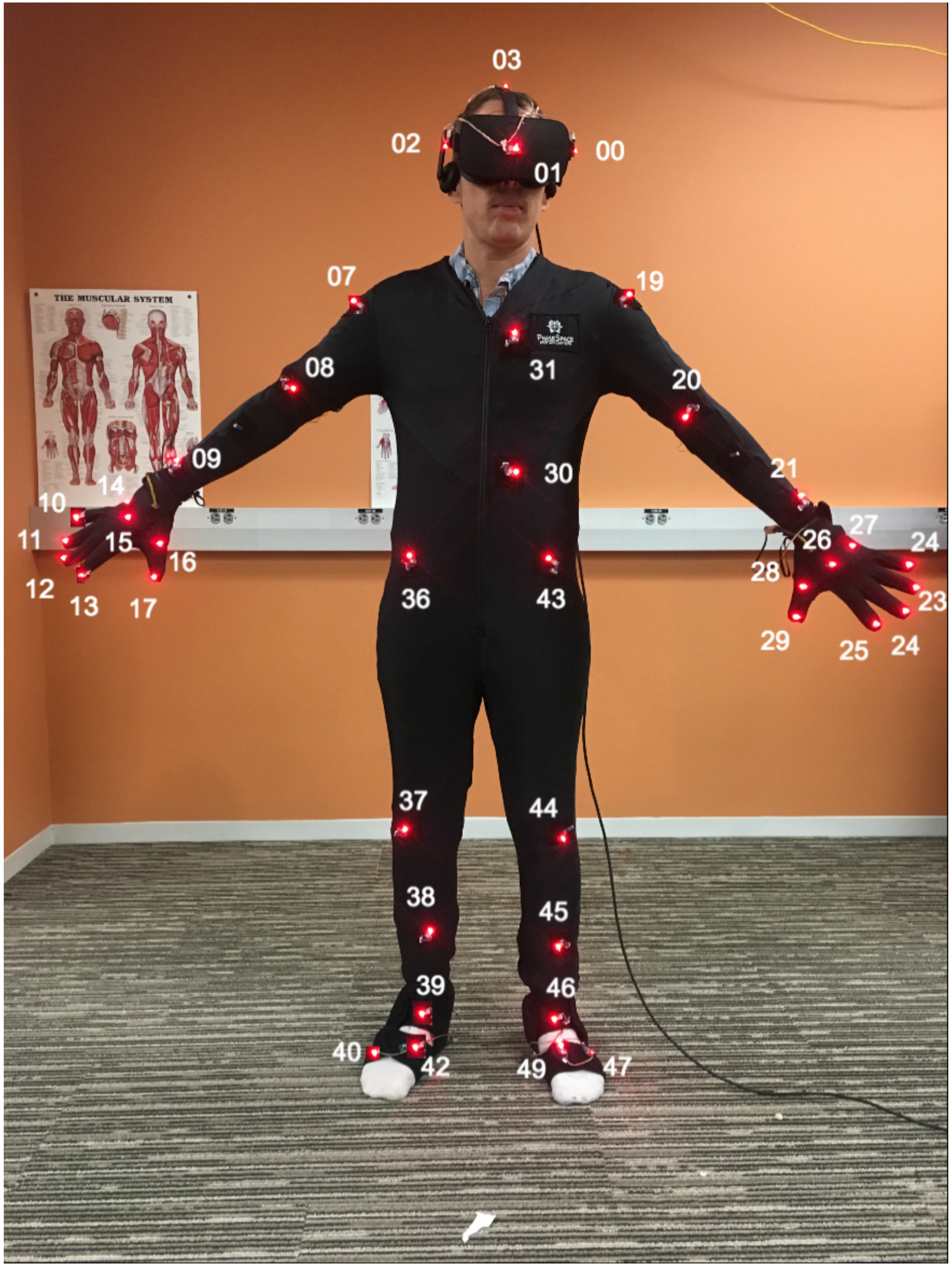
(a), $\ell = 1$



(b), $\ell = 0.3$



(c), $\ell = 3$


03
02
00
01
07
19
THE MUSCULAR SYSTEM
08
31
20
09
21
30
14
10
27
11
15
26
24
16
28
12
13
17
36
43
23
29
25
24
37
44
38
45
39
46
40
42
49
47