Bayesian Machine Learning
Course 67564
Solution To Exercise 4: Gaussian Processes

Barak Haim 0

05/01/2023

Contents

1 Theoretical

1.1 Q1

$$-\frac{1}{\alpha}\left\|x-x`\right\|^2 \overset{\alpha\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow_{p_1} \qquad \exp\left[-\frac{1}{\alpha}\left\|x-x`\right\|^2\right] \overset{\alpha\to\infty}{\longrightarrow} 1$$

$$-\alpha\left\|x-x`\right\|^2 \overset{\alpha\to\infty}{\longrightarrow} -\infty \qquad \Longrightarrow_{p_2} \qquad \exp\left[-\alpha\left\|x-x`\right\|^2\right] \overset{\alpha\to\infty}{\longrightarrow} \begin{cases} 1 & x=x` \\ 0 & else \end{cases}$$

Now, $f=(f_1,f_2)^T \sim \mathcal{N}\left(0,C_{p_i}\right)$ **and**

$$C_{p_i}=\begin{bmatrix} k_{p_i}(x_1,x_1) & k_{p_i}(x_1,x_2) \\ k_{p_i}(x_2,x_1) & k_{p_i}(x_2,x_2) \end{bmatrix}=\begin{bmatrix} 1 & k_{p_i}(x_1,x_2) \\ k_{p_i}(x_2,x_1) & 1 \end{bmatrix}$$

Specifically -

$$C_{p_1} \overset{\alpha\to\infty}{=} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 11^T$$

$$C_{p_2} \overset{\alpha\to\infty}{=} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Given by the C_{p_i}**matrices is the variance and covariance of** f_1, f_2 **namely:**

$$C_{p_i}=\begin{bmatrix} var_{p_i}[f_1] & cov_{p_i}[f_1,f_2] \\ cov_{p_i}[f_2,f_1] & var_{p_i}[f_2] \end{bmatrix}$$

So:

$$corr_{p_1}(f1,f_2)=\frac{cov_{p_1}[f_1,f_2]}{\sqrt{var_{p_1}[f_1]}\cdot\sqrt{var_{p_1}[f_2]}} \overset{\alpha\to\infty}{=} \frac{1}{\sqrt{1}\cdot\sqrt{1}}=1$$

$$corr_{p_2}(f1,f_2)=\frac{cov_{p_2}[f_1,f_2]}{\sqrt{var_{p_2}[f_1]}\cdot\sqrt{var_{p_2}[f_2]}} \overset{\alpha\to\infty}{=} \frac{0}{\sqrt{1}\cdot\sqrt{1}}=0$$

Q.E.D.

1.2 Q2

$$C_{p_2}=\begin{bmatrix} k_{p_i}(x_1,x_1) & \dots & k_{p_i}(x_1,x_N) \\ \vdots & \ddots & \vdots \\ k_{p_2}(x_N,x_1) & \dots & k_{p_2}(x_N,x_N) \end{bmatrix}=\begin{bmatrix} 1 & & \exp\left[-\alpha\Delta\right] \\ & \ddots & \\ \exp\left[-\alpha\Delta\right] & & 1 \end{bmatrix}$$

In this case we have:

$$corr_{p_2}(f1,f_2)=\frac{cov_{p_2}[f_1,f_2]}{\sqrt{var_{p_2}[f_1]}\cdot\sqrt{var_{p_2}[f_2]}} = \frac{\exp\left[-\alpha\Delta\right]}{\sqrt{1}\cdot\sqrt{1}} \approx \frac{1}{4}$$

So we have $e^{-\alpha\Delta} \approx \frac{1}{4}$ **iff** $-\alpha\Delta \approx \ln\left(\frac{1}{4}\right)$ **iff**

$$\alpha \approx -\ln\left(\frac{1}{4}\right)\cdot\frac{1}{\Delta}=\frac{\ln(4)}{\Delta}$$

Q.E.D.

1.3 Q3

a) As we saw in Q1, the correlation under $p_1$between any 2 entries of f in the limit $\alpha \to \infty$ is 1. Considering the fact that $f_i \in \{1, -1\}$, we get that (without the loss of generality) if $f_i = 1$ than $\forall j \; f_j = 1$. Else, $f_j = -1$ and so the correlation could not be 1. So, under $p_1$we get only 2 possible vectors - a vector of 1's and a vector of minus 1's and the two are equally likely so:

$$P_1(f) \overset{\alpha \to \infty}{=} \begin{cases} 0.5 & f = (1, ..., 1) \\ 0.5 & f = (-1, ..., -1) \\ 0 & else \end{cases}$$

b) Using the same logic, the correlation under $p_2$between any 2 entries of f in the limit $\alpha \to \infty$ is 0. I.e. each entry of f is completely independent of the rest. So, every value for f_i is equally possible, hence, any vector $f \in \{1, -1\}^M$ is equally possible. So:

$$P_2(f) \overset{\alpha \to \infty}{=} 2^{-M}$$

1.4 Q4

Using Bayes' law:

$$p_i(\alpha) = \frac{p_i(f)p_i(\alpha|f)}{p_i(f|\alpha)}$$

We know:

$$p_1(f_1) = p_1((1, ..., 1)|\alpha \to \infty) = 0.5$$

$$p_2(f_1) = p_2((1, ..., 1)|\alpha \to \infty) = \frac{1}{2^M}$$

Also:

$$p_1(f_2) = p_2((-1, 1, ...)|\alpha \to \infty) = 0$$

$$p_2(f_2) = p_2((-1, 1, ...)|\alpha \to \infty) = \frac{1}{2^M}$$

Hence - for each $1 \le M$ we get $p_2(f_1) \le p_1(f_1)$ and $p_1(f_2) \le p_2(f_2)$ so:

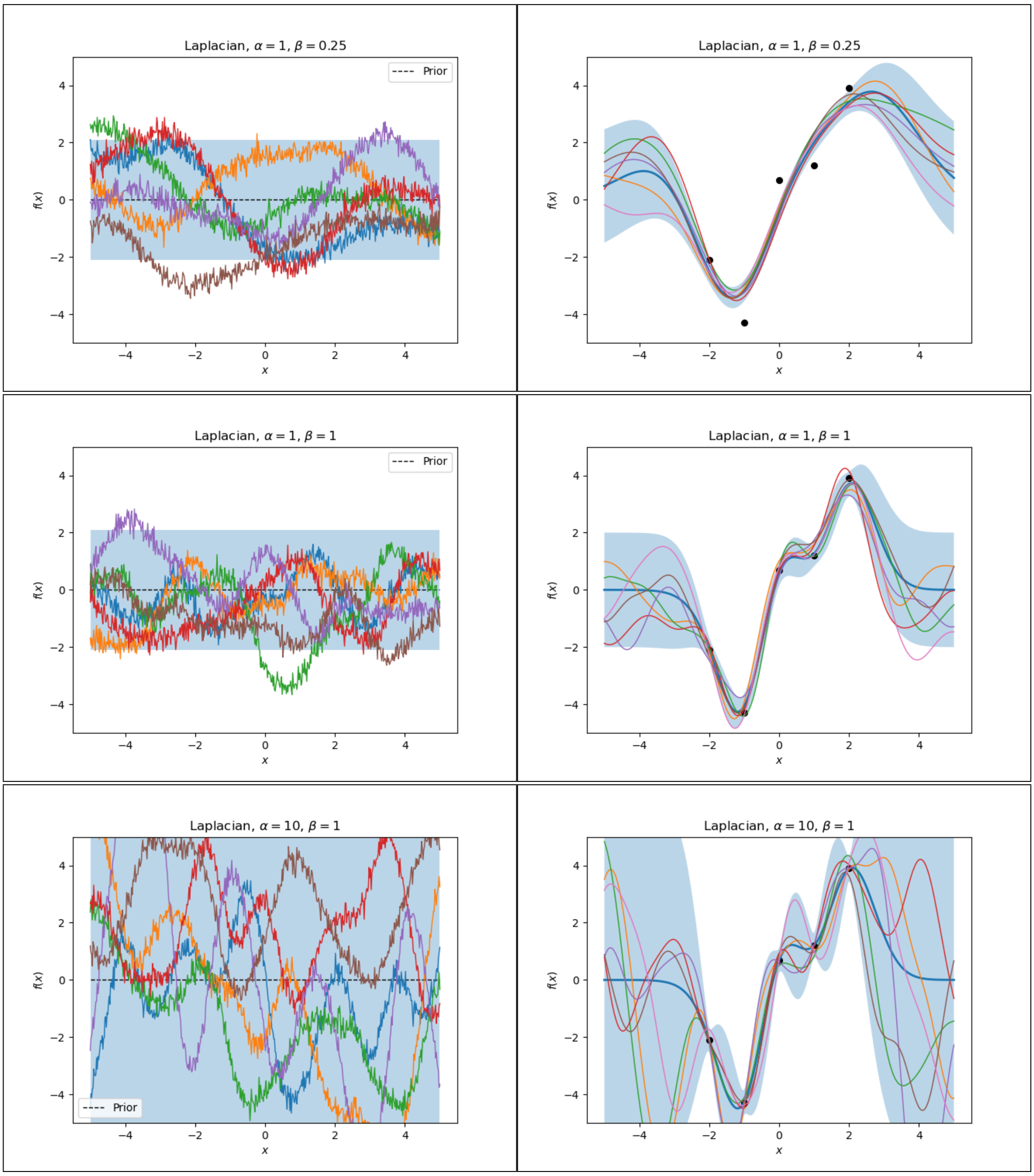
a) Given f_1- $p_1$will be selected.

b) Given f_2- $p_2$will be selected.

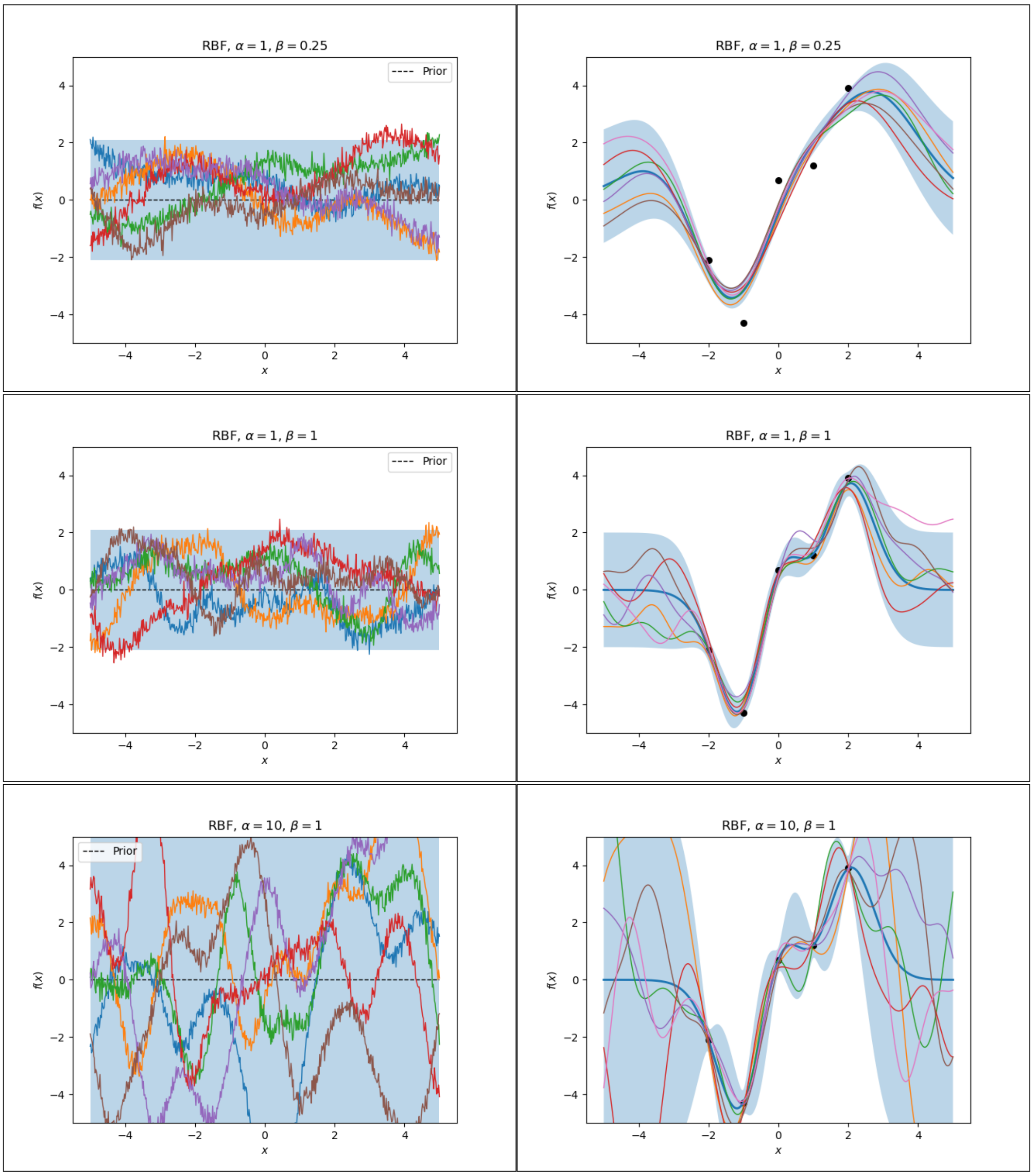
2 Practical

2.1 Kernels

2.1.1 Laplacian

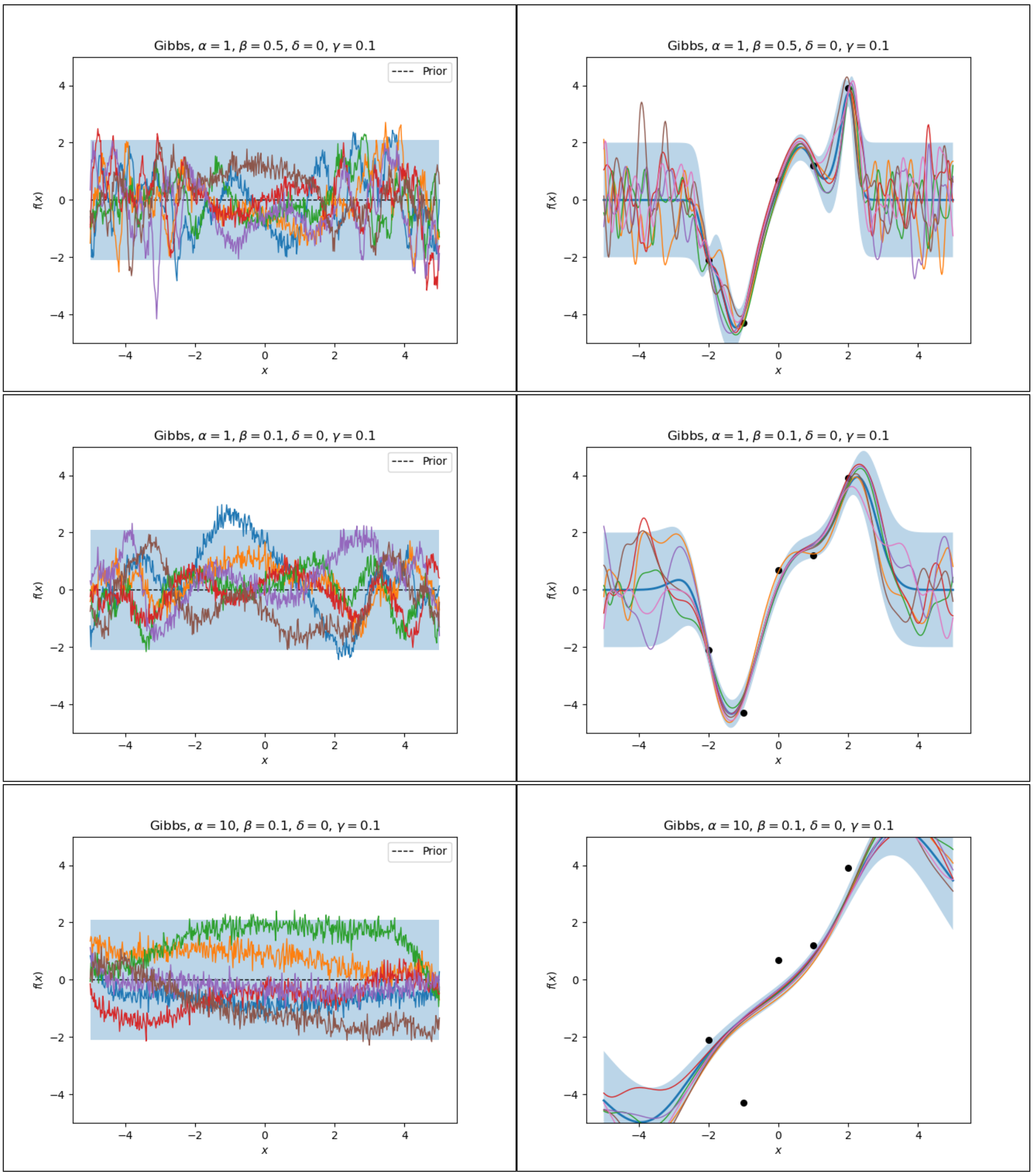


2.1.2 RBF



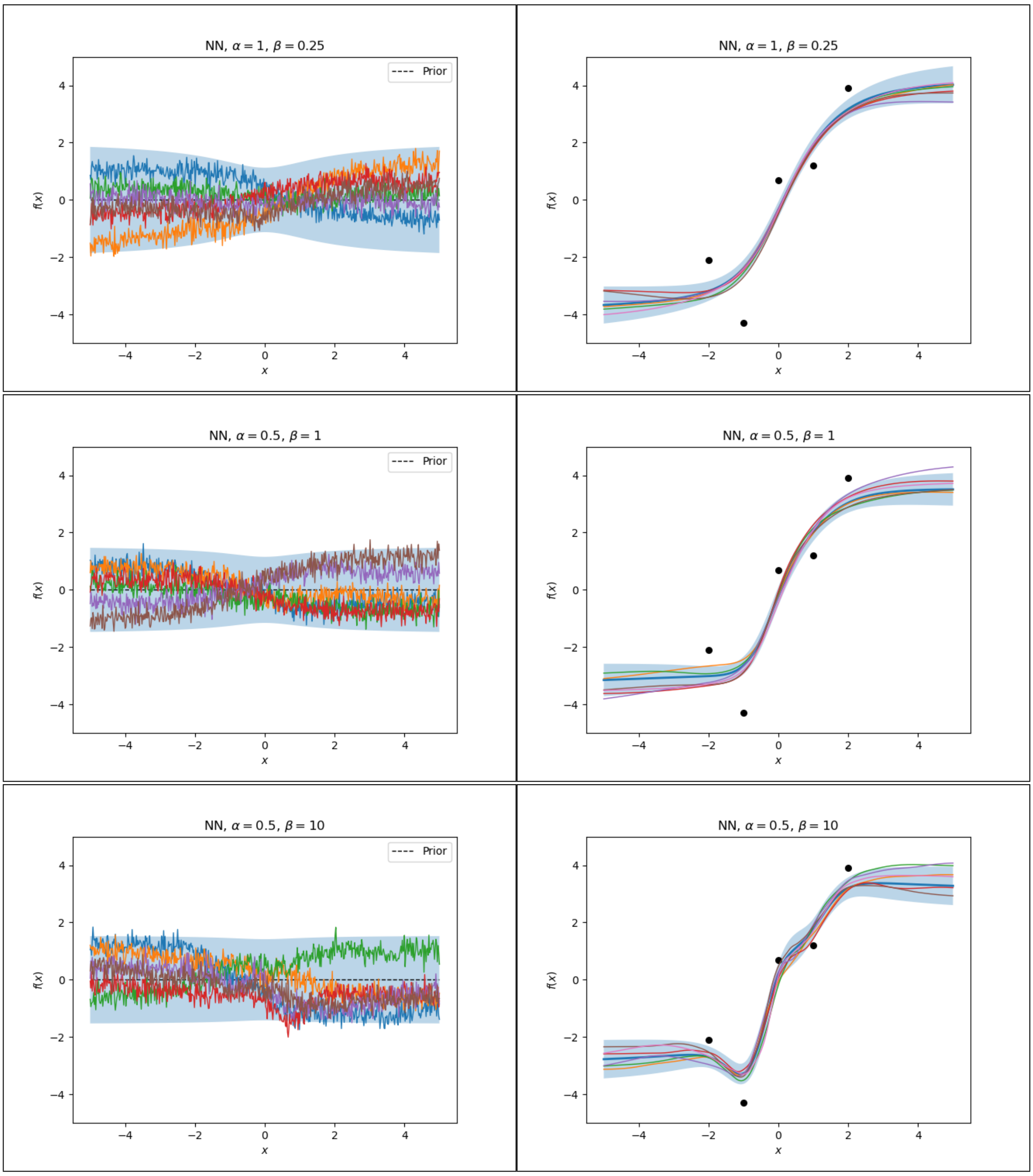
RBF, α = 1, β = 0.25
Prior
f(x)
x
RBF, α = 1, β = 0.25
f(x)
x
RBF, α = 1, β = 1
Prior
f(x)
x
RBF, α = 1, β = 1
f(x)
x
RBF, α = 10, β = 1
Prior
f(x)
x
RBF, α = 10, β = 1
f(x)
x

2.1.3 Gibbs



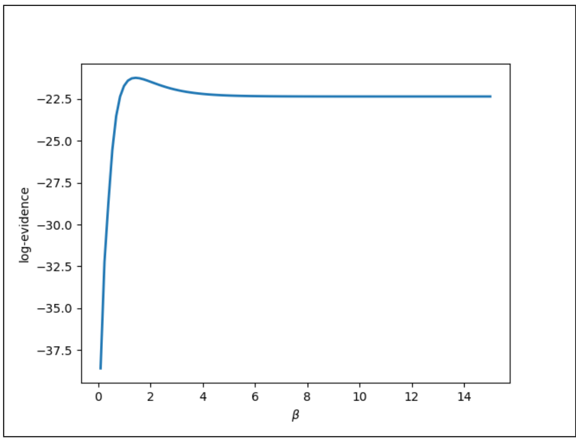
Gibbs, $\alpha = 1$, $\beta = 0.5$, $\delta = 0$, $\gamma = 0.1$
Prior
Gibbs, $\alpha = 1$, $\beta = 0.5$, $\delta = 0$, $\gamma = 0.1$
Gibbs, $\alpha = 1$, $\beta = 0.1$, $\delta = 0$, $\gamma = 0.1$
Prior
Gibbs, $\alpha = 1$, $\beta = 0.1$, $\delta = 0$, $\gamma = 0.1$
Gibbs, $\alpha = 10$, $\beta = 0.1$, $\delta = 0$, $\gamma = 0.1$
Prior
Gibbs, $\alpha = 10$, $\beta = 0.1$, $\delta = 0$, $\gamma = 0.1$
$f(x)$
x

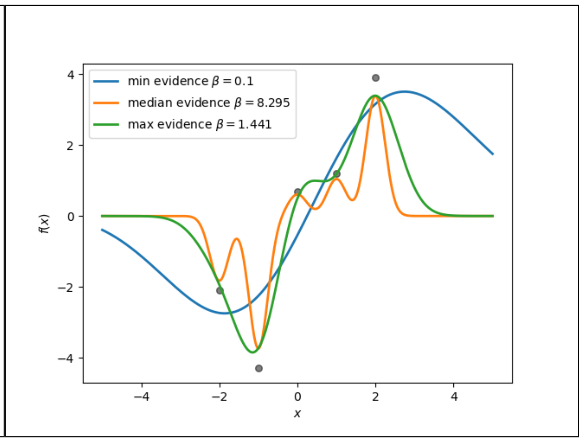
2.1.4 NN

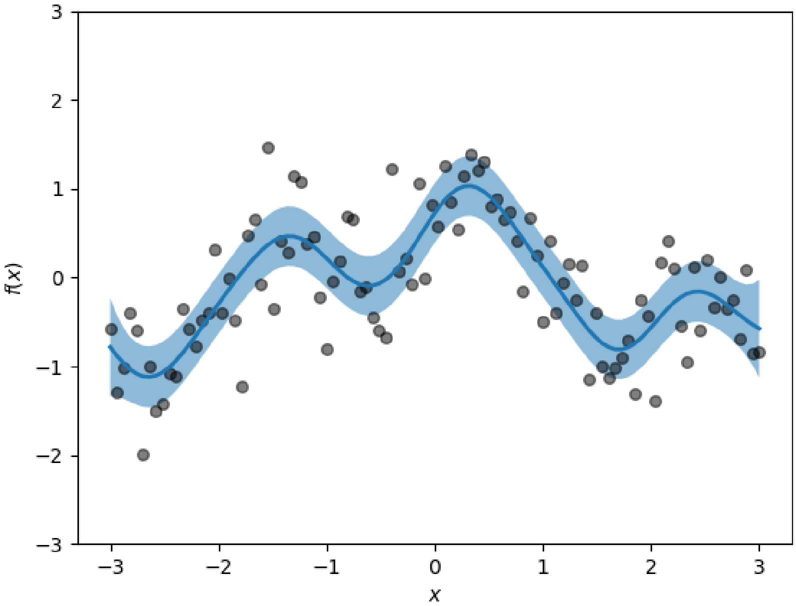


NN, $\alpha = 1$, $\beta = 0.25$
Prior
NN, $\alpha = 1$, $\beta = 0.25$
NN, $\alpha = 0.5$, $\beta = 1$
Prior
NN, $\alpha = 0.5$, $\beta = 1$
NN, $\alpha = 0.5$, $\beta = 10$
Prior
NN, $\alpha = 0.5$, $\beta = 10$
$f(x)$
x

2.2 Evidence


log-evidence
β


min evidence β = 0.1
median evidence β = 8.295
max evidence β = 1.441
f(x)
x


f(x)
x