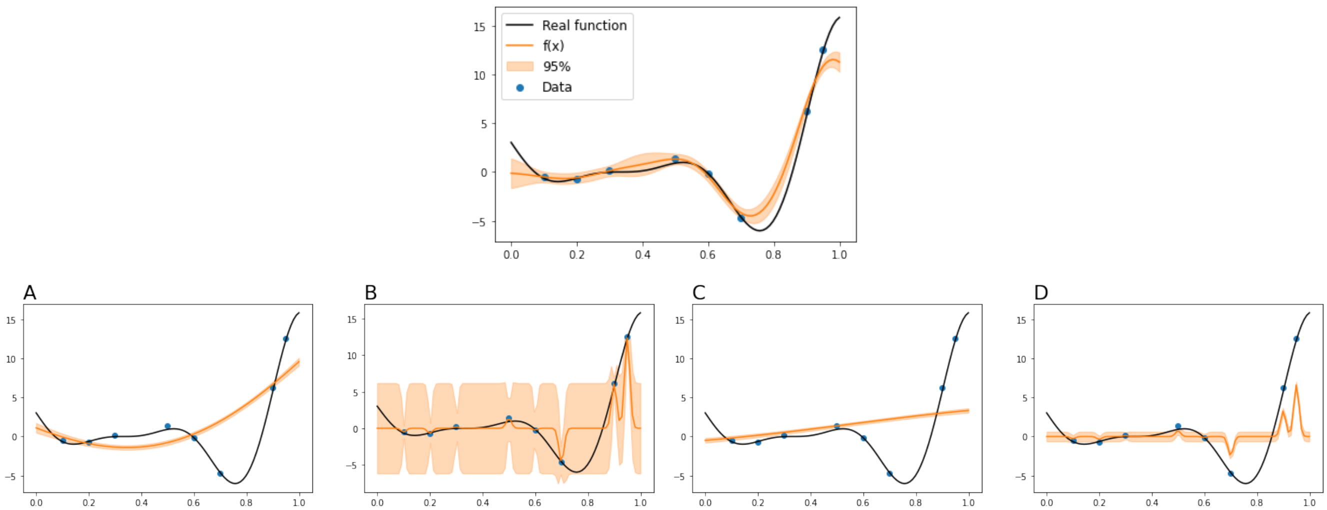
1. **Intuition Question**



(a) The top plot of the figure above shows a GP regression with the RBF kernel

$$k(x, y) = \sigma^2 \exp\left(-\frac{(x-y)^2}{2\ell^2}\right)$$

on a one-dimensional dataset. In the 4 plots below (panels A-D), the length scale ℓ and the variance σ^2 were either down- or upscaled by a factor of 10, yielding 4 combinations of the hyperparameters:

i. 0.1ℓ and $0.1\sigma^2$,

ii. 10.0ℓ and $0.1\sigma^2$,

iii. 0.1ℓ and $10.0\sigma^2$,

iv. 10.0ℓ and $10.0\sigma^2$.

Assign each of the four combinations to the corresponding plots (A-D), and provide a short justification for each assignment.

(b) **Theory Question** - This question explores the connection between the feature map view and the kernel view of regression. Consider the normal linear regression problem in the weight representation $\mathbf{w}$, where $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; \mathbf{0}, I)$, for some feature space $\mathbf{w} \in \mathbb{R}^F$, and the likelihood of some observations $\mathbf{y} \in \mathbb{R}^N$ (at locations $X \in \mathbb{R}^N$) is $p(\mathbf{y}|\mathbf{w}, X) = \mathcal{N}(\mathbf{y}; \Phi(X)\mathbf{w}, \sigma^2 I)$. The feature map is defined by the function $\Phi(x) = [1, x, , x^2, ...x^{F-1}]^\top$.

i. Using results from lectures and previous exercise sheets, write down the posterior distribution $p(\mathbf{w}|\mathbf{y}, X)$.

ii. Now, suppose we want to predict the values of the fitted function at new values X^*. By the laws of Gaussians from the lecture, calculate the mean of the predictive distribution of y^* at X^*, that is, write down the mean of $p(\mathbf{y}^*|\mathbf{y}, X, X^*)$.

iii. Now, define a kernel function by $k(x, y) = \sum_{i=1}^{F} \Phi_i(x)\Phi_i(y) = \sum_{i=0}^{F-1} x^i y^i$. Suppose we define a Gaussian Process prior over the data, $y \sim GP(0, k(x, y))$. Derive the mean of the predictive distribution over X^*, conditioned on the observations $(X, \mathbf{y})$. That is, write down the predictive mean $\mu_{y^*|y,X,X^*}$. Does this match your result for part ii? As an extension, you can also verify that the covariance matrices match.

iv. (Optional challenge) The RBF (square-exponential kernel) was defined in the lectures as $k(x, y) = \sigma^2 \exp(-\frac{(x-y)^2}{2l^2})$. We said that this corresponded to infinitely many basis functions (i.e., a feature map $\Phi(x)$ that is infinitely dimensional). Show that, to write down the kernel as a sum of basis functions as in part c, infinitely many features are required. [Hint: expand $\exp(-(x-y)^2)$ as an infinite sum].

(c) **Coding Question** - See `Exercise_04.ipynb`