COMP4702/COMP7703/DATA7703 - Machine Learning
Homework 10 - Gaussian processes

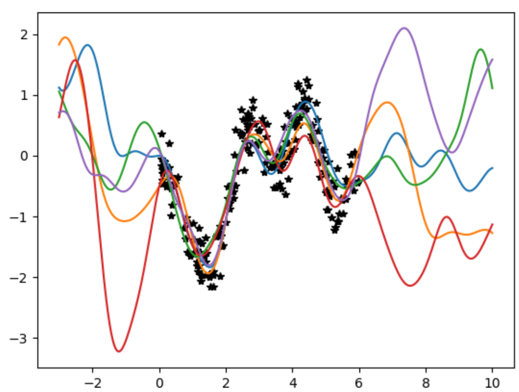
Marcus Gallagher

Core Questions

1. Using the same data as the prac and the covariance function

$$k(\mathbf{x}_p, \mathbf{x}_q) = \exp\Big(-\tfrac{1}{2\ell^2}\|\mathbf{x}_p - \mathbf{x}_q\|_2^2\Big)\exp\Big(-\tfrac{2}{\ell^2}\sin^2\big(\pi\|\mathbf{x}_p - \mathbf{x}_q\|_2/P\big)\Big),$$

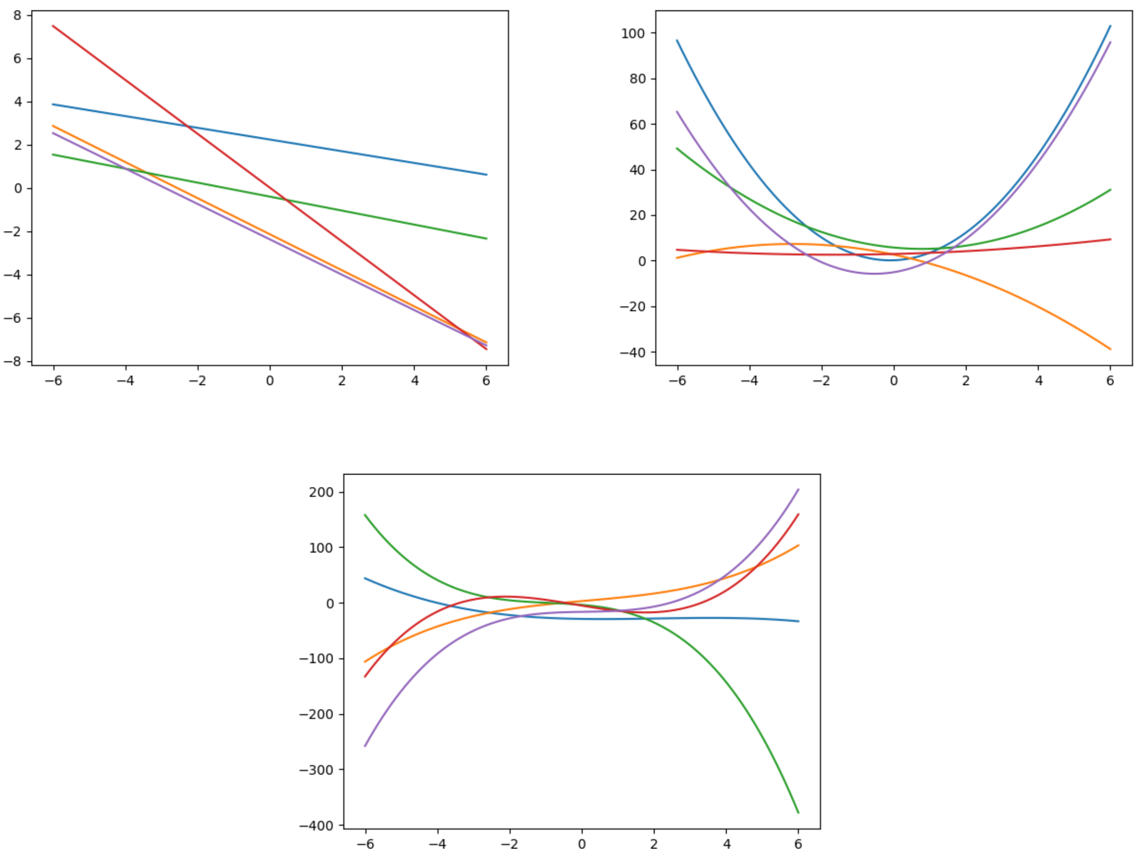
produce a plot as in Question 6 of the prac with $\sigma_n = 1$, $\ell = 2$ and $P = 3$. You may use any implementation of Gaussian process regression.
Answer: (Also label axes)



2. Consider the covariance function

$$k(\mathbf{x}_p, \mathbf{x}_q) = (\mathbf{x}_p \cdot \mathbf{x}_q + c)^d$$

For each value of $d = 1$, $d = 2$, $d = 3$, plot 5 random draws (as in the figure on page 2 of the prac) on the interval $[-6, 6]$ from a Gaussian process prior with the above kernel when $c = 5$. Use any implementation to sample from the Gaussian process prior.
Answer: (Also label axes)



3. Based on your answer to question 2 or otherwise, guess which function class the covariance function in question 2 represents.
Answer: The first plot contains linear functions. The second plot contains quadratic functions. The third plot contains cubic functons. Polynomials of order d (any mention of c is acceptable but not required)

Extension Questions

4. **Bayesian treatment of standard linear regression.** Let X be an $n \times d$ matrix of training inputs with corresponding regression targets represented in an n dimensional vector $\mathbf{y}$. Let $\mathbf{x}_i$ and y_i be the ith row and element of X and $\mathbf{y}$ respectively. Assume a model of the form.

$$\begin{aligned} y &= f(\mathbf{x}) + \epsilon \\ f(\mathbf{x}) &= \mathbf{x} \cdot \mathbf{w} \\ \epsilon &\sim \mathcal{N}(0, \sigma_n^2), \end{aligned}$$

where ϵ is independent of $\mathbf{w}$.

(a) What is the likelihood, $p(\mathbf{y}|X, \mathbf{w})$? (Hint: What is the distribution of y if $f(\mathbf{x})$ is a fixed number?)

Answer: Given $\mathbf{x}$ and $\mathbf{w}$, y is Gaussian with mean $f(\mathbf{x})$ and variance σ_n^2. $\mathbf{y}$ contains iid copies of y and is just a product of these Gaussians. Any reasonable notation or word answer of this fact is acceptable e.g.

$$p(\mathbf{y}|X,\mathbf{w}) = \frac{1}{2\pi\sqrt{\det\Sigma}}\exp\Big(-\frac{1}{2}\big(\mathbf{y}-X\mathbf{w}\big)^\top\Sigma^{-1}\big(\mathbf{y}-X\mathbf{w}\big)\Big), \quad \Sigma = \sigma_n^2 I$$

or

$$\mathbf{y}|X,\mathbf{w} \sim \mathcal{N}(X\mathbf{w}, \sigma_n^2 I),$$

or abusing notation,

$$p(\mathbf{y}|X,\mathbf{w}) = \mathcal{N}(X\mathbf{w}, \sigma_n^2 I),$$

or multivariate Gaussian with mean $X\mathbf{w}$ and covariance $\sigma_n^2 I$ etc.

(b) Put a prior over $\mathbf{w}$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, I)$, where I is the identity matrix. Show that the log posterior over $\mathbf{w}$ is given by

$$\log p(\mathbf{w}|X,\mathbf{y}) = -\frac{1}{2\sigma_n^2}\Big(\mathbf{y}-X\mathbf{w}\Big)^\top\Big(\mathbf{y}-X\mathbf{w}\Big) - \frac{1}{2}\|\mathbf{w}\|^2 + C,$$

where C is a constant not depending on $\mathbf{w}$. (Hint: start with Bayes' rule)
Answer:

$$\begin{aligned}
p(\mathbf{w}|X,\mathbf{y}) &= \frac{p(y|\mathbf{w},X)p(\mathbf{w})}{p(\mathbf{y}|X)} \\
\log p(\mathbf{w}|X,\mathbf{y}) &= \log p(y|\mathbf{w},X) + \log p(\mathbf{w}) - \log p(\mathbf{y}|X) \\
&= -\frac{1}{2\sigma_n^2}\Big(\mathbf{y}-X\mathbf{w}\Big)^\top\Big(\mathbf{y}-X\mathbf{w}\Big) - \frac{1}{2}\|\mathbf{w}\|^2 + C
\end{aligned}$$

(c) Does the maximum in $\mathbf{w}$ over the log posterior (the MAP) correspond to the solution obtained by any other machine learning technique(s) that you are aware of? Which one(s)?
Answer: Yes. e.g. penalised least squares, regularised least squares, ridge regression, ...

(d) This model (in both the prior and the posterior) is a Gaussian process. What is the covariance function of the Gaussian process prior over $f(\mathbf{x})$?
Answer: Either by inspection/intuition state the answer (e.g. we are doing a Bayesian linear regression, so linear kernel, linear covariance function, dot product, ...) or with working,

$$\begin{aligned}
\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] - \mathbb{E}[f(\mathbf{x})]\mathbb{E}[f(\mathbf{x}')] &= \mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] \\
&= \mathbb{E}[\sum_i w_i x_i \sum_j w_j x'_j] \\
&= \mathbb{E}[\sum_i w_i^2 x_i x'_i] \\
&= \sum_i x_i x'_i \\
&= \mathbf{x}\cdot\mathbf{x}'
\end{aligned}$$