Coursework 2
Mathematics for Machine Learning (70015)

Instructions

This coursework has both writing and coding components. The python code you submit must compile on a standard CSG Linux installation.

You are not permitted to use any symbolic manipulation libraries (e.g. `sympy`) or automatic differentiation tools (e.g. `tensorflow`) for your submitted code (though, of course, you may find these useful for checking your answers). Your code will be checked for imports. Please use `matplotlib` for plotting. You should not need to import anything other than `numpy` and `matplotlib` for the submitted code for this assignment.

Your coding answers you should include all code, including plotting code. Your Python script should run and output the plots in the current folder (see the `savefig` function). The report should include the plots that are requested, together with a short discussion of notable features. If you are required to derive something, include the derivation in your report.

No aspect of your submission may be hand-drawn. You are strongly encouraged to use LaTeXto create the written component.

Submit the following to CATE before the deadline:

- A file `write_up.pdf` for your written answers.
- A file `coding_answers.py` which implements all the methods for the coding exercises.

Data
The regression questions will use the same 1D data. These data are pairs (x_i, y_i) where x_i are 25 values uniformly spaced in [0, 0.9], and $y_i = g(x_i)$, where

$$g(x) = \cos(10x^2) + 0.1\sin(100x).$$

We collect the x_i in $\boldsymbol{X}$ and the y_i in $\boldsymbol{y}$.

In python, this dataset can be generated as follows:

```
import numpy as np
N = 25
X = np.reshape(np.linspace(0, 0.9, N), (N, 1))
Y = np.cos(10*X**2) + 0.1 * np.sin(100*X)
```

Basis Functions

In the regression questions, we will use the following classes of basis functions $\boldsymbol{\phi}(x) = (1, \phi_1(x), ..., \phi_J(x))^T$ where $J + 1$ is the dimension of the basis functions.

- Polynomial of degree K:
$$\phi_j(x) = x^j,$$
for $j = 1, 2, \ldots, K$.

- Trigonometric of degree K with unit frequency:
$$\phi_{2j-1}(x) = \sin(2\pi jx)$$
$$\phi_{2j}(x) = \cos(2\pi jx)$$
for $j = 1, 2, \ldots, K$

The 1 in the first position is to absorb the bias term into the weights. This simplifies the notation, but note that Kth degree polynomial basis functions are of dimension $K + 1$, Kth degree trigonometric basis functions are of dimension $2K + 1$, and the Gaussian basis functions with K means are of dimension $K + 1$. Another useful notation is the $N \times M$ design matrix, defined as $(\boldsymbol{\Phi})_{nm} = (\boldsymbol{\phi}_m(x_n))$, where $n = 1, 2, ..., N$ indexes the data points and $m = 1, 2, ..., M$ indexes the basis functions.

1 Linear Regression

In this question we consider a factorizing likelihood

$$p(\boldsymbol{y}|\boldsymbol{X}) = \prod_{i=1}^{N} p(y_i|x_i)$$

and Gaussian linear model

$$y_i \sim \mathcal{N}(\boldsymbol{w}^T \boldsymbol{\phi}(x_i), \sigma^2)$$

with basis functions as defined in the instructions section. We will often be changing the basis functions so all your derivations should be in terms of $\boldsymbol{\phi}$.

In this question we will find the maximum likelihood solution for the parameters, conditioned on the data. The data is defined in instruction section.

a) **[5 marks]** By first finding the maximum likelihood solution for the parameters σ^2 and $\boldsymbol{w}$ in terms of $\boldsymbol{\Phi}$, plot the predicted mean at test points in the interval $[-0.3, 1.3]$ in the case of polynomial basis functions of order 0, 1, 2, 3 and order 11. Plot all the curves on the same axes, showing also the data.

b) **[5 marks]** Repeat the previous part but this time with trigonometric basis functions of orders 1 and 11. Use test points in [-1, 1.2] to see the periodicity. Note that your basis functions should be of size $2J + 1$ for order J (i.e., don't forget the bias term)

c) **[6 marks]** In this part, you will investigate over-fitting with leave-one-out cross validation. You should use trigonometric basis functions of order 0 to 10 inclusive. For each choice, use leave-one-out cross-validation to estimate the average squared test error. Plot this average error on a graph against order of basis together. On the same graph plot also the maximum likelihood value for σ^2.

d) **[6 marks]** Briefly describe the concept of over-fitting, using your graph in the previous part as an illustrative example. You should also refer to your plots from the first two parts of this question. Describe why over-fitting occurs, and what can be done to prevent it.