Gaussian Process Programing Assignment

Equations follow the development in Marsland *Machine Learning* CRC Press 2014

1. **The data format**

   The data is organized by subject.Each subject traced the curve 5 times. The horizontal index indexes the position data for markers sequentially in x-y-z format. The vertifical index is time. The scale is in video frames at $\frac{1}{60}$ secpnd per frame. Sometimes the motion capture device misses data and this is usually indicated with a blank space.

2. **The learning algorithm** By choosing sample points and hyper parameters you can construct a covariance K. The goal is to pick hyper parameters that maximize the log probability of seeing the data, which is specified by

   $$\log P(\mathbf{y}|\mathbf{f},\theta) = \frac{1}{2}\mathbf{f}^T(\mathbf{K}+\sigma_n^2)^{-1}\mathbf{f} - \frac{1}{2}\log|\mathbf{K}+\sigma_n^2| - \textsf{constant}$$

   Where σ is a hyper parameter the strategy is to compute $\frac{\partial P}{\partial \sigma}$ and use gradient descent to find a local optimum. However computing these dirivative takes several steps. Examining the formula, we can see that we will need the derivatives of $\mathbf{K}^{-1}$ and $\mathbf{K}$. This prcess is fairly involved and takes several steps. Writing $\mathbf{Q} = \mathbf{K} + \sigma_n^2$ allows us to specify these derivatives as

   $$\frac{\partial \mathbf{Q}}{\partial \sigma} = -\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \sigma}\mathbf{Q}^{-1}$$

   and

   $$\frac{\partial \log|\mathbf{Q}|}{\partial \sigma} = \textsf{trace}\ \left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \sigma}\right)$$

   Almost done; just have to differentiate the kernel function containing the hyper parameters

First the kerenel:

$$
\begin{aligned}
k(\mathbf{t}, \mathbf{t}') &= \exp(\sigma_f) \exp\left(-\frac{1}{2}\exp(\sigma_l)|\mathbf{t} - \mathbf{t}'|^2\right) + \exp(\sigma_n)\mathbf{I} \\
&= k' + \exp(\sigma_n)\mathbf{I}
\end{aligned}
$$

Now, finally we go for $\frac{\partial \mathbf{Q}}{\partial \sigma}$:

$$
\begin{aligned}
\frac{\partial k}{\partial \sigma_f} &= k' \\
\frac{\partial k}{\partial \sigma_l} &= k'\left(-\frac{1}{2}\exp(\sigma_l)|\mathbf{t} - \mathbf{t}'|^2\right) \\
\frac{\partial k}{\partial \sigma_n} &= \exp(\sigma_n)\mathbf{I}
\end{aligned}
$$

All these steps have to be combined to get your derivatives.

3. **Getting things to work**

   Compared to the examples covered in lecture, the data is much more dense, so you might have to play with it to get a nice result. One way would be to subsample the data but you want to be care full not to filter out the possibly changing statistics over time. Credit-wise, you can get almost full points by juts getting the learning algorithm to work and come up with a reasonable hyper parameters estimate. For Jedi-level, you should test the hypothesis that allowing the hyperparameters to vary in time is a better model than just having a global set.

   There is too much data in the data set. A suggestion is to just pick one marker and one coordinate and fit that, extending your analyses if that works out. Good luck.