Derivation of some Gaussian Process Expressions from GP Class

John Shawe-Taylor

November, 2017

1 Introduction

This note contains a more detailed derivation of some of the expressions in the notes.

2 Posterior Distribution

First request was for a derivation of the form of the posterior distribution on slide 30

$$\begin{aligned}
P_{\text{post}}(\mathbf{w}|S) &\propto \exp\left(-\frac{\sum_{i=1}^{m}(y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle)^2}{2\sigma^2} - \frac{\|\mathbf{w}\|^2}{2}\right) \\
&\propto \exp\left(-\frac{-2y_i'\mathbf{X}\mathbf{w} + \mathbf{w}'\mathbf{X}'\mathbf{X}\mathbf{w}}{2\sigma^2} - \frac{\|\mathbf{w}\|^2}{2}\right)
\end{aligned}$$

the expression in brackets is a quadratic form that is negative definite and hence defines a Gaussian distribution. To identify the mean and covariance matrix we need to transform the expression into the standard form. We can easily identify the covariance matrix by looking at the quadratic terms in $\mathbf{w}$:

$$-\frac{1}{2}\mathbf{w}'\left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} + I\right)\mathbf{w}.$$

This implies that the covariance matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2}(\mathbf{X}'\mathbf{X} + \sigma^2\mathbf{I})$$

We also know the mean of the distribution is $\mathbf{w}_{\text{map}}$, the map solution. Hence, we have

$$
\begin{aligned}
P_{\text{post}}(\mathbf{w}|S) &\propto \exp\left(-\frac{\sum_{i=1}^m (y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle)^2}{2\sigma^2} - \frac{\|\mathbf{w}\|^2}{2}\right) \\
&\propto \exp\left(-\frac{-2y_i'\mathbf{X}\mathbf{w} + \mathbf{w}'\mathbf{X}'\mathbf{X}\mathbf{w}}{2\sigma^2} - \frac{\|\mathbf{w}\|^2}{2}\right) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{w} - \mathbf{w}_{\text{map}})'\mathbf{\Sigma}^{-1}(\mathbf{w} - \mathbf{w}_{\text{map}})\right).
\end{aligned}
$$

i.e. a Gaussian with mean $\mathbf{w}_{\text{map}}$ and covariance $\mathbf{\Sigma}$.

3 Error bars

The second request was for the derivation of the error bar expression on page 31. First consider the following product:

$$
\begin{aligned}
&(\mathbf{I} + a\mathbf{X}'\mathbf{X})(\mathbf{I} - a\mathbf{X}'(\mathbf{I} + a\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}) \\
&\quad = \mathbf{I} + a\mathbf{X}'\mathbf{X} - a\mathbf{X}'(\mathbf{I} + a\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} - a^2\mathbf{X}'\mathbf{X}\mathbf{X}'(\mathbf{I} + a\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} \\
&\quad = \mathbf{I} + a\mathbf{X}'\mathbf{X} - a\mathbf{X}'(\mathbf{I} + a\mathbf{X}\mathbf{X}')(\mathbf{I} + a\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} \\
&\quad = \mathbf{I} + a\mathbf{X}'\mathbf{X} - a\mathbf{X}'\mathbf{X} = \mathbf{I}
\end{aligned}
$$

Hence,

$$
(\mathbf{I} + a\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{I} - a\mathbf{X}'(\mathbf{I} + a\mathbf{X}\mathbf{X}')^{-1}\mathbf{X})
$$

Using this equality in the expression $\phi(\mathbf{x})'\mathbf{\Sigma}\phi(\mathbf{x})$ with $a = \sigma^{-2}$ gives

$$
\begin{aligned}
\phi(\mathbf{x})'\mathbf{\Sigma}\phi(\mathbf{x}) &= \phi(\mathbf{x})'\mathbf{\Sigma}\phi(\mathbf{x}) \\
&= \phi(\mathbf{x})'(\sigma^{-2}\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\phi(\mathbf{x}) \\
&= \phi(\mathbf{x})'(\mathbf{I} - \sigma^{-2}\mathbf{X}'(\mathbf{I} + \sigma^{-2}\mathbf{X}\mathbf{X}')^{-1}\mathbf{X})\phi(\mathbf{x}) \\
&= \kappa(\mathbf{x}, \mathbf{x}) - \sigma^{-2}\mathbf{k}'(\mathbf{I} + \sigma^{-2}\mathbf{C})^{-1}\mathbf{k} \\
&= \kappa(\mathbf{x}, \mathbf{x}) - \mathbf{k}'(\mathbf{C} + \sigma^2\mathbf{I})^{-1}\mathbf{k}
\end{aligned}
$$

as in the notes.

4 Evidence

The third request was for a derivation of the expression for the evidence on page 32.

Here we consider the Gaussian process view of the prior and posterior distribution. The prior distribution over the outputs $\mathbf{z}$ given the inputs $S_{\mathbf{x}}$ is given by

$$P(\mathbf{z}|S_{\mathbf{x}}) - \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{C}^{-1}\mathbf{z}\right)$$

where $\mathbf{C}$ is the covariance matrix on $S_{\mathbf{x}}$. Similarly, the noise probability is now

$$P(\mathbf{y}|\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{z}\|^2\right).$$

Hence, the probability of the data in the model is given by

$$\begin{aligned}
P(\mathbf{y}|S_x) &= \int P(\mathbf{z}|S_{\mathbf{x}})P(\mathbf{y}|\mathbf{z})d\mathbf{z} \\
&= \int \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{C}^{-1}\mathbf{z}\right) \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{z}\|^2\right) d\mathbf{z} \\
&= \frac{1}{\sqrt{\det(2\pi\mathbf{C})}(2\pi\sigma^2)^{m/2}} \int \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{C}^{-1}\mathbf{z} - \frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{z}\|^2\right) d\mathbf{z}.
\end{aligned}$$

We now need to identify the Gaussian distribution defined by the expression in brackets again by identifying the quadratic terms, this time in $\mathbf{z}$:

$$-\frac{1}{2}\mathbf{z}'\left(\mathbf{C}^{-1} + \frac{1}{\sigma^2}\mathbf{I}\right)\mathbf{z} =: -\frac{1}{2}\mathbf{z}'\mathbf{\Sigma}^{-1}\mathbf{z}.$$

Now completing the square we have

$$\begin{aligned}
-\frac{1}{2}\mathbf{z}'\mathbf{C}^{-1}\mathbf{z} &- \frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{z}\|^2 \\
&= -\frac{1}{2}(\mathbf{z}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{z}-\mu) - \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y} + \frac{1}{2}\mu'\Sigma^{-1}\mu
\end{aligned}$$

We can also compare the linear terms in $\mathbf{z}$ to obtain

$$\frac{1}{\sigma^2}\mathbf{y} = \mathbf{\Sigma}^{-1}\mu. \text{ or } \mu = \mathbf{C}(\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{y}$$

agreeing with the dual Ridge Regression solution. Hence, we can evaluate the integral as

$$\begin{aligned}
P(\mathbf{y}|S_{\mathbf{x}}) &= \frac{1}{\sqrt{\det(2\pi\mathbf{C})}(2\pi\sigma^2)^{m/2}} \sqrt{\det(2\pi\mathbf{\Sigma})} \exp\left(-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y} + \frac{1}{2}\mu'\Sigma^{-1}\mu\right) \\
&= \frac{1}{\sqrt{\det(\mathbf{C}\mathbf{\Sigma}^{-1}\sigma^2)}(2\pi)^{m/2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y} + \frac{1}{2}\mu'\Sigma^{-1}\mu\right) \\
&= \frac{1}{\sqrt{\det(\mathbf{C}+\sigma^2\mathbf{I})}(2\pi)^{m/2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y} + \frac{1}{2}\mu'\Sigma^{-1}\mu\right)
\end{aligned}$$

Finally, using the solution of μ the expression in brackets can be simplified:

$$\begin{aligned}
-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y} &+ \frac{1}{2}\mu'\Sigma^{-1}\mu \\
&= \frac{1}{2\sigma^2}\mathbf{y}'\left((\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{C}(\sigma^2\mathbf{C}^{-1}+\mathbf{I})\mathbf{C}(\mathbf{C}+\sigma^2\mathbf{I})^{-1}-\mathbf{I}\right)\mathbf{y} \\
&= \frac{1}{2\sigma^2}\mathbf{y}'\left((\mathbf{C}+\sigma^2\mathbf{I})^{-1}(\sigma^2\mathbf{I}+\mathbf{C})\mathbf{C}(\mathbf{C}+\sigma^2\mathbf{I})^{-1}-\mathbf{I}\right)\mathbf{y} \\
&= \frac{1}{2\sigma^2}\mathbf{y}'\left(\mathbf{C}(\mathbf{C}+\sigma^2\mathbf{I})^{-1}-\mathbf{I}\right)\mathbf{y} \\
&= \frac{1}{2\sigma^2}\mathbf{y}'\left(\mathbf{C}-(\mathbf{C}+\sigma^2\mathbf{I})\right)(\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{y} \\
&= -\frac{1}{2}\mathbf{y}'(\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{y}
\end{aligned}$$

So we have the final expression

$$\log(P(\mathbf{y}|S_{\mathbf{x}})) = -\frac{1}{2}\mathbf{y}'(\mathbf{C}+\sigma^2\mathbf{I})^{-1}\mathbf{y} - \frac{m}{2}\log(2\pi) - \frac{1}{2}\log\det(\mathbf{C}+\sigma^2\mathbf{I}),$$

as in the notes.