Machine Learning 1
Home Work 2

Arthur Bražinskas
University of Amsterdam

September 18, 2015

1 Posterior predictive distributions (3.2)

1.1

First of all it's important to note that t_* is scalar and therefore the likelihood is a univariate Gaussian density function, and there is no need for t_* transpose.

1.2

$$p(t_*|\underline{\phi_*}, \underline{w}, \beta)p(\underline{w}|\underline{m_N}, \underline{\underline{S_N}}) = \mathcal{N}(t_*|\underline{\phi_*}^T\underline{w}, \frac{1}{\beta})\mathcal{N}(\underline{w}|\underline{m_N}, \underline{\underline{S_N}}) =$$

$$= \sqrt{\frac{\beta}{2\pi}}\exp(-\frac{\beta}{2}(t_*-\underline{\phi_*^T}\underline{w})^T(t_*-\underline{\phi_*^T}\underline{w}))\frac{|\underline{\underline{S_N}}|^{-\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}}\exp(-\frac{1}{2}(\underline{w}-\underline{m_N})^T\underline{\underline{S_N}}^{-1}(\underline{w}-\underline{m_N}))$$

Now we set $C_L = \sqrt{\frac{\beta}{2\pi}}$ to be the constant of the likelihood, and $C_P = \frac{|\underline{\underline{S_N}}|^{-\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}}$ to be our posterior constant.

$$= C_L C_p \exp(-\frac{\beta}{2}(t_* - \underline{\phi_*^T}\underline{w})^T(t_* - \underline{\phi_*^T}\underline{w}))\exp(-\frac{1}{2}(\underline{w} - \underline{m_N})^T\underline{\underline{S_N}}^{-1}(\underline{w} - \underline{m_N})) =$$

Next we expand our exponent terms.

$$= C_L C_p \exp(-\frac{1}{2}(\beta t_* t_* - \beta t_*\underline{\phi_*^T}\underline{w} - \beta t_*\underline{\phi_*^T}\underline{w} + \beta\underline{\phi_*^T}\underline{w}\underline{\phi_*^T}\underline{w} + \underline{w^T}\underline{\underline{S_N^{-1}}}\underline{w} - \underline{w^T}\underline{\underline{S_N^{-1}}}\underline{m_N} - \underline{m_N}^T\underline{\underline{S_N}}^{-1}\underline{w} + \underline{m_N^T}\underline{\underline{S_N^{-1}}}\underline{m_N}))$$

Now we are going to collect all the terms that depend on w in one exponent, and the rest in another one. And we simplify our clauses.

$$= C_L C_p \exp(-\frac{1}{2}(\underline{w}^T(\beta\underline{\phi_*^T}\underline{\phi_*} + S_N^{-1})\underline{w} - 2\underline{w}^T(\beta\underline{\phi_*^T}t_* + S_N^{-1}m_N)))\exp(-\frac{1}{2}(\beta t_* t_* + m_N^T S_N^{-1}m_N))$$

In the next step we set $S_{N+1}^{-1} = \beta\underline{\phi}_*^T\underline{\phi}_* + S_N^{-1}$ and $\underline{\underline{S}}_{N+1}^{-1} m_{N+1} = \beta\underline{\phi}_* t_* + S_N^{-1} m_N$ and rewrite our expression accordingly.

$$= C_L C_p \exp(-\frac{1}{2}(\underline{w}^T S_{N+1}^{-1}\underline{w} - 2w^T S_{N+1}^{-1})) \exp(-\frac{1}{2}(\beta t_* t_* + m_N^T S_N^{-1} m_N))$$

1.3

Next we complete the square for w by adding $m_{N+1}^T S_{N+1}^{-1} m_{N+1}$ to the exponent with w terms and subtract from the second exponent.

$$= C_L C_p \exp(-\frac{1}{2}(\underline{w}^T S_{N+1}^{-1}\underline{w} - 2w^T S_{N+1}^{-1} + m_{N+1}^T S_{N+1}^{-1} m_{N+1})) \exp(-\frac{1}{2}(\beta t_* t_* + m_N^T S_N^{-1} m_N - m_{N+1}^T S_{N+1}^{-1} m_{N+1}))$$

As we will need m_{N+1} and $m_{N+1}^T S_{N+1}^{-1} m_{N+1}$ in the future, we going to solve for it now.

$$\underline{\underline{S}}_{N+1}^{-1} m_{N+1} = \beta\underline{\phi}_*^T t_* + S_N^{-1} m_N$$

$$m_{N+1} = S_{N+1}(\beta\underline{\phi}_* t_* + S_N^{-1} m_N) = (S_{N+1}^{-1} = \beta\underline{\phi}_*^T\underline{\phi}_* + S_N^{-1})^{-1}(\beta\underline{\phi}_* t_* + S_N^{-1} m_N)$$

$$m_{N+1}^T S_{N+1}^{-1} m_{N+1} = (\beta\underline{\phi}_* t_* + S_N^{-1} m_N)^T S_{N+1}(\beta\underline{\phi}_* t_* + S_N^{-1} m_N)$$

1.4

After a bit of simplifications we obtain the w exponential in a Gaussian form:

$$p(t_*|\underline{\phi_*}, \underline{w}, \beta) p(\underline{w}|\underline{m_N}, \underline{\underline{S_N}}) =$$

$$= C_L C_p \exp(-\frac{1}{2}[(\underline{w} - m_{N+1})^T S_{N+1}^{-1}(\underline{w} - m_{N+1})]) \exp(-\frac{1}{2}(m_N^T S_N^{-1} m_N + \beta t_* t_* - m_{N+1}^T S_{N+1}^{-1} m_{N+1}))$$

So now as we have a joint of likelihood and posterior we can integrate over w.

$$p(t_*|\underline{\phi}_*, \underline{\underline{\phi}}, \underline{t}, \alpha, \beta) = \int p(t_*|\underline{\phi_*}, \underline{w}, \beta) p(\underline{w}|\underline{m_N}, \underline{\underline{S_N}}) dw =$$

Now we set $\frac{1}{C_w} = \int \exp(-\frac{1}{2}[(\underline{w} - m_{N+1})^T S_{N+1}^{-1}(\underline{w} - m_{N+1})])dw$, and as the second exponent does not depend on w it remains present the equation outside of the integral.

$$= \frac{C_L C_p}{C_w} \exp(-\frac{1}{2}(m_N^T S_N^{-1} m_N + \beta t_* t_* - m_{N+1}^T S_{N+1}^{-1} m_{N+1}))$$

1.5

What we obtain in the exponent is not in a Gaussian form, so we will expand the $m_{N+1}^T S_{N+1}^{-1} m_{N+1}$ term see sub-section 1.3, and will collect linear and squared terms of t_* as a a first step of obtaining the form we want.

$$m_N^T S_N^{-1} m_N + \beta t_* t_* - m_{N+1}^T S_{N+1}^{-1} m_{N+1} =$$

$$m_N^T S_N^{-1} m_N + \beta t_* t_* - (\beta \underline{\phi}_* t_* + S_N^{-1} m_N)^T S_{N+1} (\beta \underline{\phi}_* t_* + S_N^{-1} m_N) =$$

$$= m_N^T S_N^{-1} m_N + \beta t_* t_* - m_{N+1}^T S_{N+1}^{-1} m_{N+1} =$$

$$= m_N^T S_N^{-1} m_N + \beta t_* t_* - \beta\beta t_* t_* \underline{\phi}_*^T S_{N+1} \underline{\phi}_* - \beta t_* \underline{\phi}_* S_{N+1} S_N^{-1} m_N - \beta t_* \underline{\phi}_* S_{N+1} S_N^{-1} m_N - S_N^{-1} m_N S_{N+1} S_N^{-1} m_N =$$

Now we collect the terms of t_*. Remember that t_* is scalar and not a vector.

$$t_*^2 (\beta \underline{\underline{I}} - \beta^2 \underline{\phi}_* S_{N+1} \underline{\phi}_*^T) - 2 t_* (\beta \underline{\phi}_* S_{N+1} S_N^{-1} m_N) + m_N^T S_N^{-1} m_N - S_N^{-1} m_N S_{N+1} S_N^{-1} m_N =$$

$$= t_*^2 \beta_* - 2 t_* \beta_* y_* + y_*^T \beta_* y_*$$

Where
$\beta_* = \beta \underline{\underline{I}} - \beta^2 \underline{\phi}_* S_{N+1} \underline{\phi}_*^T$
$\beta_* y_* = \beta \underline{\phi}_* S_{N+1} S_N^{-1} m_N$
$y_*^T \beta_* y_* = m_N^T S_N^{-1} m_N - S_N^{-1} m_N S_{N+1} S_N^{-1} m_N$

1.6

Now we are going to solve for β_* by expanding S_{N+1} and by using Woodbury identity.

$$\beta_* = \beta \underline{\underline{I}} - \beta^2 \underline{\phi}_* S_{N+1} \phi_*^T = \beta \underline{\underline{I}} - \beta \underline{\underline{I}} \underline{\phi}_* (\beta \underline{\phi}_*^T \underline{\phi}_* + S_N^{-1})^{-1} \underline{\phi}_*^T \beta \underline{\underline{I}}$$

$$= \beta \underline{\underline{I}} - \beta \underline{\underline{I}} \underline{\phi}_* (S_N^{-1} + \underline{\phi}_*^T \beta \underline{\underline{I}} \underline{\phi}_*)^{-1} \underline{\phi}_*^T (\beta \underline{\underline{I}}) = (\frac{1}{\beta} + \underline{\phi}_* S_N \underline{\phi}_*^T)^{-1}$$

1.7

Now we are going to solve for y_* by expanding S_{N+1} and β_*, and by doing some simplifications .

$$\beta_* y_* = \beta \underline{\phi}_* S_{N+1} S_N^{-1} m_N$$

$$y_* = \beta_*^{-1} \beta \underline{\phi}_* (\beta \underline{\phi}_*^T \underline{\phi}_* + S_N^{-1})^{-1} S_N^{-1} m_N =$$

$$(\frac{1}{\beta} + \underline{\phi}_* S_N \underline{\phi}_*^T) \beta \underline{\phi}_* (S_N^{-1} + \beta \underline{\phi}_*^T \underline{\phi}_*)^{-1} S_N^{-1} m_N)$$

$$= \underline{\phi}_* [(\underline{\underline{I}} + \beta S_N \underline{\phi}_*^T \underline{\phi}_*)(S_N^{-1} + \beta \underline{\phi}_*^T \underline{\phi}_*)^{-1} S_N^{-1}] m_N =$$

$$= \underline{\phi}_* [S_N S_N^{-1} (\underline{\underline{I}} + \beta S_N \underline{\phi}_*^T \underline{\phi}_*)(S_N^{-1} + \beta \underline{\phi}_*^T \underline{\phi}_*)^{-1} S_N^{-1}] m_N =$$

$$= \underline{\phi}_* [S_N (S_N^{-1} + \beta \underline{\phi}_*^T \underline{\phi}_*)(S_N^{-1} + \beta \underline{\phi}_*^T \underline{\phi}_*)^{-1} S_N^{-1}] m_N =$$

Now using the rule of matrix multiplication by it's inverse we simplify everything an obtain:

$$= \underline{\phi}_* m_N$$

1.8

Since we obtained solutions for y_* and β_* we can plug them into the Gaussian form of the posterior predictive distribution.

$$p(t_* | \underline{\phi}_*, \underline{t}, \underline{\underline{\Phi}}, \alpha, \beta) = \mathcal{N}(t_* | y_*, \frac{1}{\beta_*}) = \mathcal{N}(t_* | \underline{\phi}_* m_N, \frac{1}{\beta} + \underline{\phi}_*^T S_N \underline{\phi}_*)$$