COM6509/4509 — Tutorial Sheet 1
Bayesian and Maximum Likelihood Manipulation of Gaussian Models

Neil Lawrence

October 5th, 2012

1. Univariate Gaussian model. A Gaussian density governs a vector of univariate observations, $\mathbf{t} = \{t_i\}_{i=1}^N$. The associated error function has the following form.

$$E(\mu) = \sum_{i=1}^{N} (t_i - \mu)^2$$

   (a) Introduce the variance parameter, σ^2 and convert the error function to the Gaussian density. Find the maximum likelihood solutions for both μ and σ^2.

   (b) Place the following Gaussian prior over the mean,

$$p(\mu) = \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2\alpha}\mu^2\right)$$

   and compute the marginal likelihood for $\mathbf{t}$ and the posterior density for μ.

2. Maximum likelihood in a multivariate Gaussian. A data set consists of p dimensional vectors, $\mathbf{t}_{i,:}$ from a matrix $\mathbf{T} = \{\mathbf{t}_{i,:}\}_{i=1}^N$ (i.e. $\mathbf{T} \in \Re^{N\times p}$). The likelihood is given by

$$p(\mathbf{T}) = \prod_{i=1}^{N} p(\mathbf{t}_{i,:})$$

   where the likelihood of each data point is

$$p(\mathbf{t}_{i,:}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{t}_{i,:} - \boldsymbol{\mu})^\top \mathbf{C}^{-1} (\mathbf{t}_{i,:} - \boldsymbol{\mu})\right).$$

   (a) Write down the log likelihood and use the following matrix and vector derivatives

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{A}^\top \mathbf{x} \\
\frac{\mathrm{d}\log|\mathbf{C}|}{\mathrm{d}\mathbf{C}} &= \mathbf{C}^{-1} \\
\frac{\mathrm{d}\mathbf{a}^\top \mathbf{C}^{-1}\mathbf{a}}{\mathrm{d}\mathbf{C}} &= -\mathbf{C}^{-1}\mathbf{a}\mathbf{a}^\top \mathbf{C}^{-1}
\end{aligned}$$

to show that the maximum likelihood solutions for the mean, $\hat{\boldsymbol{\mu}}$ and covariance matrix, $\hat{\mathbf{C}}$, are

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{t}_{i,:},$$
$$\hat{\mathbf{C}} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{t}_{i,:} - \hat{\boldsymbol{\mu}})(\mathbf{t}_{i,:} - \hat{\boldsymbol{\mu}})^\top.$$

(b) Now consider an independent Gaussian prior over the elements of the mean vector,

$$p(\boldsymbol{\mu}) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2\alpha}\mu_i^2\right)$$

i. Show that this can be written in vector form as follows:

$$p(\boldsymbol{\mu}) = \frac{1}{(2\pi\alpha)^{\frac{p}{2}}} \exp\left(-\frac{1}{2\alpha}\boldsymbol{\mu}^\top\boldsymbol{\mu}\right).$$

ii. Now compute the posterior density for $\boldsymbol{\mu}$, $p(\boldsymbol{\mu}|\mathbf{T})$. Write down the terms that remain that would be required for the marginal likelihood of $\mathbf{T}$, $p(\mathbf{T})$ (note given the matrix algebra we've covered you won't be able to write down the full form of the marginal likelihood).

3. **Regression with a basis function model**. Assume that we wish to perform a nonlinear regression by computing a set of basis functions, for example,

$$\phi_j(\mathbf{x}_{i,:}) = \exp\left(-\frac{1}{2\ell_j^2}(x_i - \mu_j)^2\right),$$

where μ is a location parameter and ℓ is a width parameter for the jth basis function. For each data point we take the m basis functions and write them in a vector of the following form

$$\boldsymbol{\phi}_{i,:} = [\phi_1(\mathbf{x}_{i,:}) \ldots \phi_m(\mathbf{x}_{i,:})]^\top$$

and the complete set of basis functions is written in a matrix, $\boldsymbol{\Phi} \in \Re^{N\times m}$ of the following form,

$$\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_{1,:}\boldsymbol{\phi}_{2,:}\ldots\boldsymbol{\phi}_{N,:}\right]^\top.$$

If we assume Gaussian noise we can write down the Gaussian likelihood of a single data point, i,

$$p(t_i|\boldsymbol{\phi}_{i,:},\mathbf{w}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(t_i - \mathbf{w}^\top\boldsymbol{\phi}_{i,:})^2\right).$$

(a) Assume the noise is independent and identically distributed and write down the corresponding likelihood and log likelihood of the entire data set.

(b) Show that the maximum likelihood solution for $\mathbf{w}$ is given by

$$\hat{\mathbf{w}} = \left(\mathbf{\Phi}^\top \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^\top \mathbf{t}.$$

(c) Consider a Gaussian prior over the parameters, $\mathbf{w}$,

$$p(\mathbf{w}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2\alpha} w_i^2\right).$$

Show that the posterior for $\mathbf{w}$ is given by a Gaussian with covariance

$$\mathbf{C}_w = \left(\frac{1}{\sigma^2} \mathbf{\Phi}^\top \mathbf{\Phi} + \alpha^{-1} \mathbf{I}\right)^{-1}$$

and mean

$$\boldsymbol{\mu}_w = \frac{1}{\sigma^2} \mathbf{C}_w \mathbf{\Phi}^\top \mathbf{t}$$

i. Compare the solution for the maximum likelihood and the posterior mean over $\mathbf{w}$. When do they become the same?
ii. What problems occur for the maximum likelihood solution if $m > N$?

(d) Show that the marginal likelihood of the data set is given by

$$p(\mathbf{t}|\mathbf{X}, \alpha, \sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{K}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{t}^\top \mathbf{K}^{-1} \mathbf{t}\right)$$

where

$$\mathbf{K} = \alpha \mathbf{\Phi} \mathbf{\Phi}^\top + \sigma^2 \mathbf{I}$$

by using the matrix inversion formula:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \mathbf{D}\mathbf{A}^{-1}.$$

Tutorial Sheet 1 Answers

1 a)
$$E(\mu) = \sum_{i=1}^{N} (t_i - \mu)^2$$

$$P(\underline{t} | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(t_i - \mu)^2}{2\sigma^2}\right)$$

Log likelihood

$$\log P(t | \mu, \sigma^2) = -\frac{N}{2} \log \sigma^2 - \frac{N}{2} \log 2\pi - \sum_{i=1}^{N} \frac{(t_i - \mu)^2}{2\sigma^2}$$

$$\frac{d \log P(t | \mu, \sigma^2)}{d\sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{(t_i - \mu)^2}{2\sigma^4}$$

Set to zero to find fixed point equation

$$\frac{N}{2\sigma^2} = \sum_{i=1}^{N} \frac{(t_i - \mu)^2}{2\sigma^4}$$

Multiply both sides by $\frac{2\sigma^4}{N}$

$$\sigma^2 = \sum_{i=1}^{N} \frac{(t_i - \mu)^2}{N}$$

$$\frac{d \log p(t|\mu,\sigma^2)}{d\mu} = \sum_{i=1}^{N} (t_i - \mu)$$

$$= \sum_{i=1}^{N} t_i - N\mu$$

$$\Rightarrow \mu = \sum_{i=1}^{N} t_i \Big/ N$$

1 b)

$$p(t|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(t_i-\mu)^2}{2\sigma^2}\right)$$

$$p(\mu) = \frac{1}{(2\pi\alpha)^{1/2}} \exp\left(-\frac{\mu^2}{2\alpha}\right)$$

$$p(\mu, t) = \frac{1}{(2\pi\sigma^2)^{N/2}} \frac{1}{(2\pi\alpha)^{1/2}} \exp\left(-\sum_{i=1}^{N} t_i^2 \Big/ 2\sigma^2 + \sum_{i=1}^{N} t_i\mu \Big/ \sigma^2 - \frac{N\mu^2}{2\sigma^2} - \frac{\mu^2}{2\alpha}\right)$$

Focus on the exponent

$$-\frac{\sum_{i=1}^{N} t_i^2}{2\sigma^2} + \frac{\sum_{i=1}^{N} t_i \mu}{\sigma^2} - \frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\mu^2$$

Complete the square to find posterior for μ

variance must be $\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)^{-1}$ to match quadratic term in μ.

What is the mean $(\bar{\mu})$ required to match linear term in μ?

$$-\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)(\mu - \bar{\mu})^2 = -\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\mu^2$$

$$+ \left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\bar{\mu}\mu - \frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\bar{\mu}^2$$

to match this term to above

$$\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\bar{\mu} = \sum_{i=1}^{N} t_i \Big/ \sigma^2$$

POSTERIOR

Which implies $\bar{\mu} = \left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)^{-1} \sigma^{-2} \sum_{i=1}^{N} t_i$

$$p(\mu | t) = \frac{1}{\left(2\pi\left(\frac{N}{\sigma^2}+\frac{1}{\alpha}\right)^{-1}\right)^{1/2}} \exp\left(\frac{-(t - \bar{\mu})^2}{2\left(\frac{N}{\sigma^2}+\frac{1}{\alpha}\right)^{-1}}\right).$$

The remaining terms in the quadratic form that are unaccounted for are these, are from marginal likelihood

$$\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\bar{\mu}^2 - \frac{1}{2}\sum_{i=1}^{N} t_i^2 \Big/ \sigma^2$$

This term was generated to allow us to complete the square

This was a term constant in μ in original form

$$p(\mu, t) = p(\underline{t}|\mu)\, p(\mu) = p(\mu|\underline{t})\, p(\underline{t})$$

the terms in exponent
for this posterior
are given in quadratic
form. That leaves

$$\bar{\mu} = \left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)^{-1} \sigma^{-2} \sum_{i=1}^{N} t_i$$

$$\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)\bar{\mu}^2 = \sigma^{-4}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)^{-1}\left(\sum_{i=1}^{N} t_i\right)\sum_{i=1}^{N} t_i$$

Use $1^T t = \sum_{i=1}^{N} t_i$

vector of ones

$$= \sigma^{-4}\left(\frac{N}{\sigma^2} + \frac{1}{\alpha}\right)^{-1} t^T 1 1^T t$$

$$\frac{1}{2\sigma^2}\sum t_i^2 = \frac{1}{2\sigma^2} t^T t$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N} t_i^2 + \left(\frac{N}{\sigma^2}+\frac{1}{\alpha}\right)^{-1}\sigma^{-4}\left(\sum t_i\right)^2$$

$$= -\frac{1}{2} t^T \underbrace{\left[I\sigma^{-2} - \sigma^{-4}\left(\frac{N}{\sigma^2}+\frac{1}{\alpha}\right) 1 1^T\right]}_{} t$$

This is inverse covariance, covariance is

$$C_t = \left[I\sigma^{-2} - \sigma^{-4}\left(\frac{N}{\sigma^2}+\frac{1}{\alpha}\right)^{-1} 1 1^T\right]^{-1}$$

Use Matrix inversion Lemma

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}$$

$$A = \sigma^2 I \qquad B = 1 \qquad D = 1^T \qquad C = \alpha$$

$$\Rightarrow C_t = I\sigma^2 + \alpha 1 1^T$$

$$p(\underline{t}) = \frac{1}{(2\pi)^{N/2} \left| I\sigma^2 + \alpha \mathbf{1}\mathbf{1}^T \right|^{1/2}} \exp\left(-\frac{1}{2} t^T \left[I\sigma^2 + \alpha \mathbf{1}\mathbf{1}^T\right] t\right)$$

2a) $$p(t_i) = \frac{1}{(2\pi)^{P/2} |C|^{1/2}} \exp\left(-\frac{1}{2}(t_{i,:}-\mu)^T C^{-1} (t_{i,:}-\mu)\right)$$

$$p(T) = \prod_{i=1}^{N} p(t_{i,:})$$

$$= \frac{1}{(2\pi)^{Np/2} |C|^{N/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} (t_{i,:}-\mu)^T C^{-1} (t_{i,:}-\mu)\right)$$

$$\log p(T) = -\frac{Np}{2} \log 2\pi - \frac{N}{2} \log |C| - \frac{1}{2} \sum_{i=1}^{N} (t_{i,:}-\mu)^T C^{-1} (t_{i,:}-\mu)$$

$$\frac{d \log p(T)}{d\mu} = -\frac{1}{2} \sum_{i=1}^{N} \frac{d\, \underline{(t_{i,:}-\mu)^T C^{-1}} (t_{i,:}-\mu)}{d\mu}$$

$$= \sum_{i=1}^{N} C^{-1} (t_{i,:}-\mu)$$

$$= C^{-1} \sum_{i=1}^{N} t_{i,:} - N C^{-1} \mu$$

Fixed point is gradient zero

$$\Rightarrow N C^{-1} \mu = C^{-1} \sum_{i=1}^{N} t_{i,:}$$

$$\Rightarrow \mu = \frac{\sum_{i=1}^{N} t_{i,:}}{N}$$

$$\frac{d \log p(T)}{dC} = -\frac{1}{2} \frac{d \log |C|}{dC}$$

$$-\frac{1}{2} \sum_{i=1}^{N} \frac{d (t_{i,:} - \mu)^T C^{-1} (t_{i,:} - \mu)}{dC}$$

$$= -\frac{1}{2} C^{-1} + \frac{1}{2} C^{-1} \sum_{i=1}^{N} (t_i - \mu)(t_i - \mu)^T C^{-1}$$

Find fixed point by setting to zero

$$\frac{1}{2}C^{-1} = \frac{1}{2}C^{-1}\sum_{i=1}^{N}(t_i-\mu)(t_i-\mu)^T C^{-1}$$

pre multiply by $2C$ & post multiply by C to give

$$\underline{\underline{C = \sum_{i=1}^{N}(t_i-\mu)(t_i-\mu)^T}}$$

2b)(i) $p(\mu) = \prod_{i=1}^{p} \frac{1}{(2\pi\alpha)^{1/2}} \exp\left(-\frac{1}{2\alpha}\mu_i^2\right)$

$$= \frac{1}{(2\pi\alpha)^{p/2}} \exp\left(-\frac{1}{2\alpha}\underbrace{\sum_{i=1}^{p}\mu_i^2}_{=\mu^T\mu}\right)$$

$$= \frac{1}{(2\pi\alpha)^{p/2}} \exp\left(-\frac{1}{2\alpha}\mu^T\mu\right)$$

2b(ii)

$$p(t|\mu) = \frac{1}{(2\pi)^{\frac{Np}{2}} |C|^{\frac{N}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}(t_{i,:}-\mu)^T C^{-1}(t_{i,:}-\mu)\right)$$

$$p(t,\mu) = \frac{1}{(2\pi)^{\frac{Np}{2}} |C|^{\frac{N}{2}} (2\pi\alpha)^{\frac{p}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} t_{i,:}^T C^{-1} t_{i,:} + \sum_{i=1}^{N} t_{i,:}^T C^{-1}\mu - \frac{1}{2}\mu^T\left[NC^{-1}+\alpha^{-1}I\right]\mu\right)$$

Focussing on exponent only

$$-\frac{1}{2}\sum t_{i,:}^T C^{-1} t_{i,:} + \sum_{i=1}^{N} t_{i,:}^T C^{-1}\mu - \frac{1}{2}\mu^T\left[NC^{-1}+\alpha^{-1}I\right]\mu$$

For posterior and marginal

$$p(T,\mu) = p(T|\mu)\,p(\mu) = p(\mu|T)\,p(T)$$

Extract terms in μ only to find Gaussian form for $p(\mu|T)$. This means that posterior covariance is $\left[NC^{-1}+\alpha^{-1}I\right]^{-1} = \Sigma_\mu$

Quadratic form for Gaussian posterior is

$$-\frac{1}{2}(\mu - \bar{\mu})^T \Sigma_\mu^{-1} (\mu - \bar{\mu})$$

Linear term is $\bar{\mu}^T \Sigma_\mu^{-1} \mu = \sum t_{i,:}^T C^{-1} \mu$

$$\Rightarrow \bar{\mu}^T \Sigma_\mu^{-1} = \sum t_{i,:}^T C^{-1}$$

$$\Rightarrow \bar{\mu}^T = \sum t_{i,:}^T C^{-1} \Sigma_\mu$$

POSTERIOR

$$\Rightarrow \bar{\mu} = \Sigma_\mu C^{-1} \sum t_{i,:}$$

$$p(\mu | T) = \frac{1}{(2\pi)^{D/2} |\Sigma_\mu|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mu - \bar{\mu}) \Sigma_\mu^{-1} (\mu - \bar{\mu})\right)$$

For marginal the following terms remain

$$\frac{1}{2} \bar{\mu}^T \Sigma_\mu^{-1} \bar{\mu} - \frac{1}{2} \sum_{i=1}^{N} t_{i,:}^T C^{-1} t_{i,:}$$

where $\bar{\mu} = \Sigma_\mu C^{-1} \sum_{i=1}^{N} t_{i,:}$

$$-\frac{1}{2}\left[\sum t_{i,:}^T C^{-1} t_{i,:} - \sum_{i=1}^{N} t_{i,:}^T C^{-1} \Sigma_\mu C^{-1} \sum_{i=1}^{N} t_{i,:}\right]$$

THIS FAR IS FINE GIVEN THE MATERIAL WE COVER IN THE COURSE. TO GO FURTHER YOU NEED SOME MORE ADVANCED MATRIX ALGEBRA

$$p(T) \propto \exp\left(-\frac{1}{2}\sum t_{i,:}^T C^{-1} t_{i,:} - \sum t_{i,:}^T C^{-1} \Sigma_\mu C^{-1} \sum t_{i,:}\right)$$

3a) $p(t_i \mid w, \sigma^2, x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t_i - w^T\phi(x_i))^2}{2\sigma^2}\right)$

$$p(t \mid w, \sigma^2, X) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(t_i - w^T\phi(x_i))^2}{2\sigma^2}\right)$$

$$\log p(t \mid w, \sigma^2, X) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \sigma^2 - \sum_{i=1}^{N} \frac{(t_i - w^T\phi(x_i))^2}{2\sigma^2}$$

3b) $$\frac{d \log p(t \mid w, \sigma^2, X)}{dw} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N} \frac{d\,(t_i - w^T\phi(x_i))^2}{dw}$$

$$(t_i - w^T\phi(x_i))^2 = t_i^2 - 2t_i\phi(x_i)^T w + w^T\phi(x_i)\phi(x_i)^T w$$

$$\frac{d\,\underline{\quad\quad}}{dw} = -2t_i\phi(x_i) + 2\phi(x_i)\phi(x_i)^T w$$

$$\frac{d \log p(t \mid w, \sigma^2, X)}{dw} = \frac{1}{\sigma^2}\underbrace{\sum_{i=1}^{N} t_i\phi(x_i)}_{\Phi^T t} - \frac{1}{\sigma^2}\underbrace{\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^T w}_{\Phi^T\Phi}$$

$$= \frac{1}{\sigma^2}\Phi^T t - \frac{1}{\sigma^2}\Phi^T\Phi w$$

Set to zero to find optimal w

$$\frac{1}{\sigma^2}\Phi^T\Phi w = \frac{1}{\sigma^2}\Phi^T t$$

$$\Rightarrow \quad w = \left(\Phi^T\Phi\right)^{-1}\Phi^T t$$

3c) $$p(w) = \frac{1}{(2\pi\alpha)^{M/2}} \exp\left(-\frac{1}{2\alpha} w^T w\right)$$

$$p(t, w) = \frac{1}{(2\pi\sigma^2)^{N/2}} \frac{1}{(2\pi\alpha)^{M/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} \frac{\left(t_i - w^T\phi(x_i)\right)^2}{\sigma^2} - \frac{1}{2}\frac{w^T w}{\alpha}\right)$$

Exponent is

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N} t_i^2 + \frac{1}{\sigma^2} w^T \sum_{i=1}^{N} \phi(x_i)\, t_i \underbrace{- \frac{1}{2\sigma^2} w^T \sum \phi_i(x)\phi(x_i)^T w - \frac{1}{2\alpha} w^T w}$$

$$-\frac{1}{2\sigma^2} t^T t + \frac{1}{\sigma^2} w^T \Phi^T t - \frac{1}{2} w^T \left[\frac{1}{\sigma^2} \Phi^T \Phi + \frac{1}{\alpha} I \right] w$$

Posterior form w. Covariance must be

$$\Sigma_w = \left[\frac{1}{\sigma^2} \Phi^T \Phi + \frac{1}{\alpha} I \right]^{-1}$$

$$-\frac{1}{2} (w - \mu_w)^T \Sigma_w^{-1} (w - \mu_w)$$ is form which implies

$$w^T \Sigma_w^{-1} \mu_w = \frac{1}{\sigma^2} w^T \Phi^T t$$

which implies $\mu_w = \frac{\Sigma_w \Phi^T t}{\sigma^2}$

$$p(w \mid t, x, \sigma^2) = \frac{1}{(2\pi)^{M/2} |\Sigma_w|^{-1/2}} \exp\left(-(w - \mu_w)^T \Sigma_w^{-1} (w - \mu_w) \right)$$

3d(i)

$$\mu_w = \frac{\Sigma_w}{\sigma^2} \Phi^T t \qquad \hat{w} = \left(\Phi^T \Phi\right)^{-1} \Phi^T t$$

$$\Sigma_w = \left(\frac{\Phi^T \Phi}{\sigma^2} + \frac{1}{\alpha} I\right)^{-1}$$

$$\frac{\Sigma_w}{\sigma^2} = \left(\Phi^T \Phi + \sigma^2/\alpha\, I\right)^{-1}$$

If $\frac{\sigma^2}{\alpha} \to 0$ because $\sigma^2 \to 0$ (no noise)

or $\alpha \to \infty$ (infinite variance prior)

then $\mu_w = \hat{w}$ and the solutions coincide

3c(ii) If $M > N$ then $\Phi^T \Phi$ is not full rank and $\left(\Phi^T \Phi\right)^{-1}$ is not computable. This isn't a problem for the Bayesian solution because you invert $\left(\Phi^T \Phi + \sigma^2/\alpha\, I\right)^{-1}$ and $\frac{\sigma^2}{\alpha} I$ forces the matrix to be full rank.

16

3d) Remaining terms are

From completing the square

$$-\frac{1}{2\sigma^2} t^T t + \frac{1}{2} \mu_w^T \Sigma_w^{-1} \mu_w$$

$$= -\frac{1}{2} \left[\frac{t^T t}{\sigma^2} - \frac{t^T \Phi \Sigma_w \Phi^T t}{\sigma^4} \right]$$

$$= -\frac{1}{2} t^T \underbrace{\left[\sigma^{-2} I - \sigma^{-4} \Phi \left(\alpha^{-1} I + \sigma^{-2} \Phi^T \Phi \right)^{-1} \Phi^T \right]}_{K^{-1}} t$$

Matrix inversion lemma

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B \left(C^{-1} + D A^{-1} B \right)^{-1} D A^{-1}$$

$$A = \sigma^2 I \quad B = \Phi \quad C = \alpha I$$

Gives

$$= -\frac{1}{2} t^T \underbrace{\left[\sigma^2 I + \alpha \Phi \Phi^T \right]}_{K}^{-1} t$$

$$\Rightarrow p(t) = \frac{1}{(2\pi)^{N/2} K^{1/2}} \exp\left(-\frac{1}{2} t^T K^{-1} t\right)$$

$$K = \sigma^2 I + \alpha \underline{\Phi}\,\underline{\Phi}^T$$