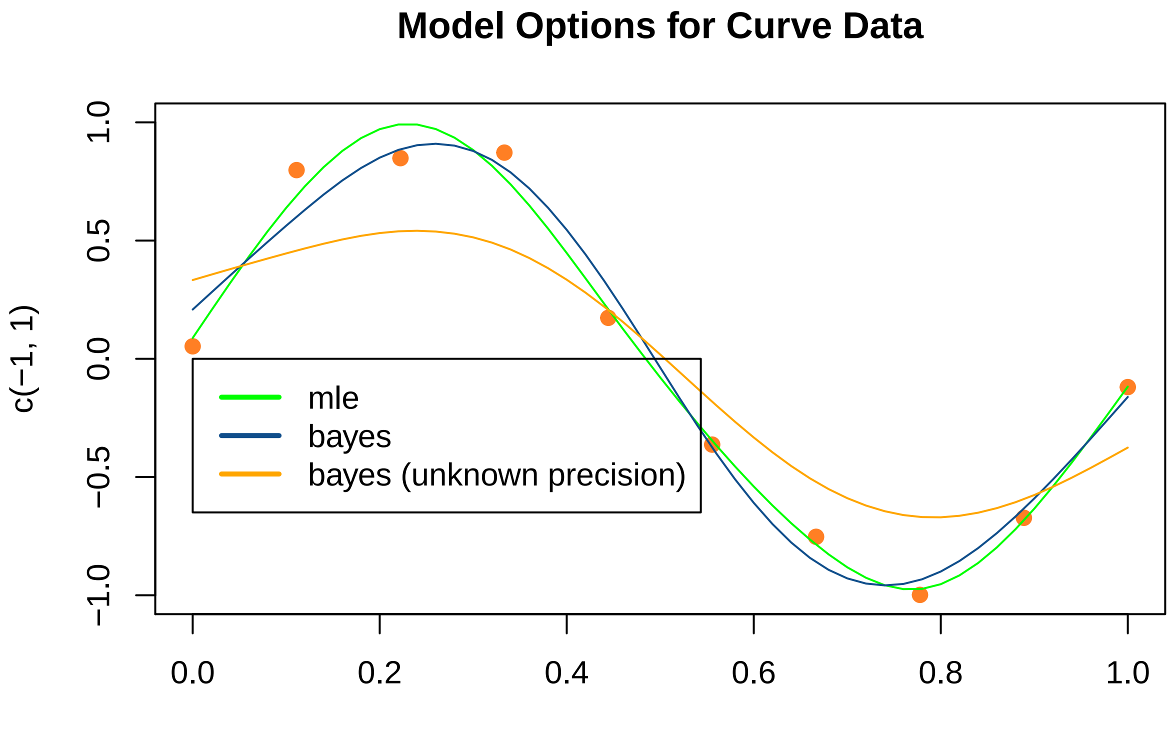
Statistical Modelling and Inference: Week 3 Prior Modelling

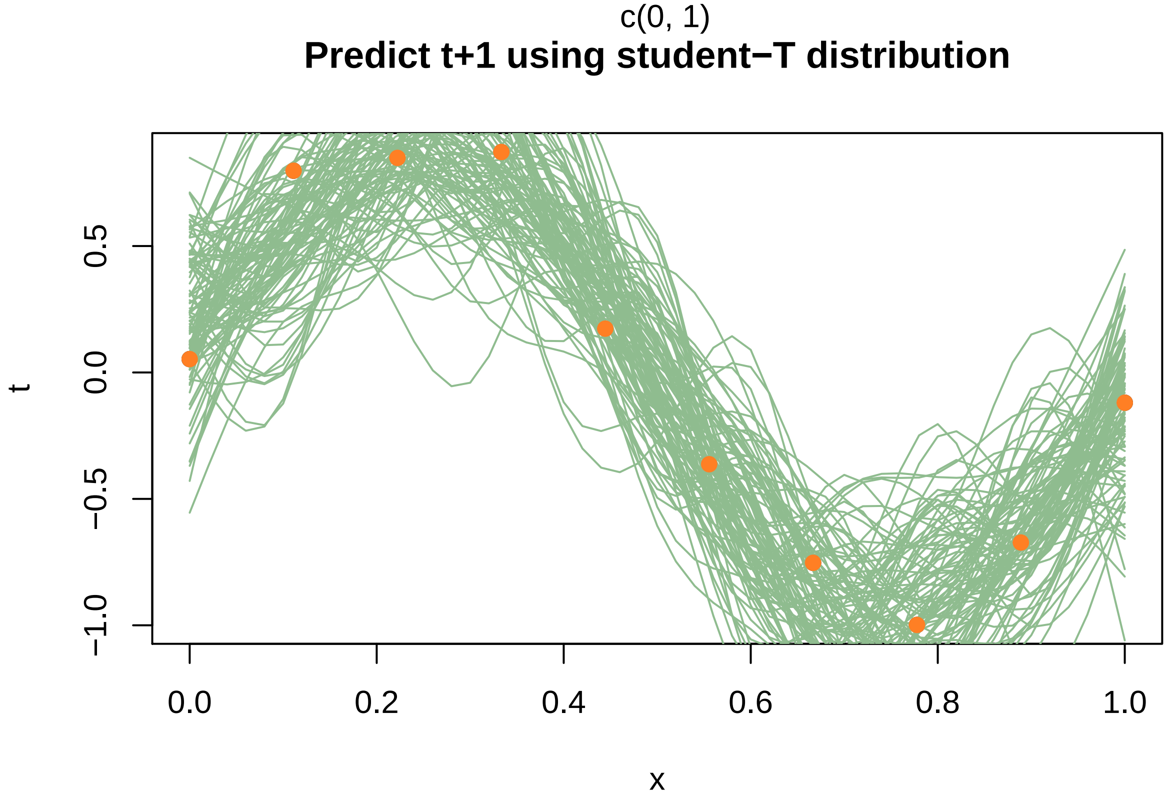
Prior Modelling ** 1. Extend your program to also learn q from the data. Work with the Normal.Gamma prior with specification**

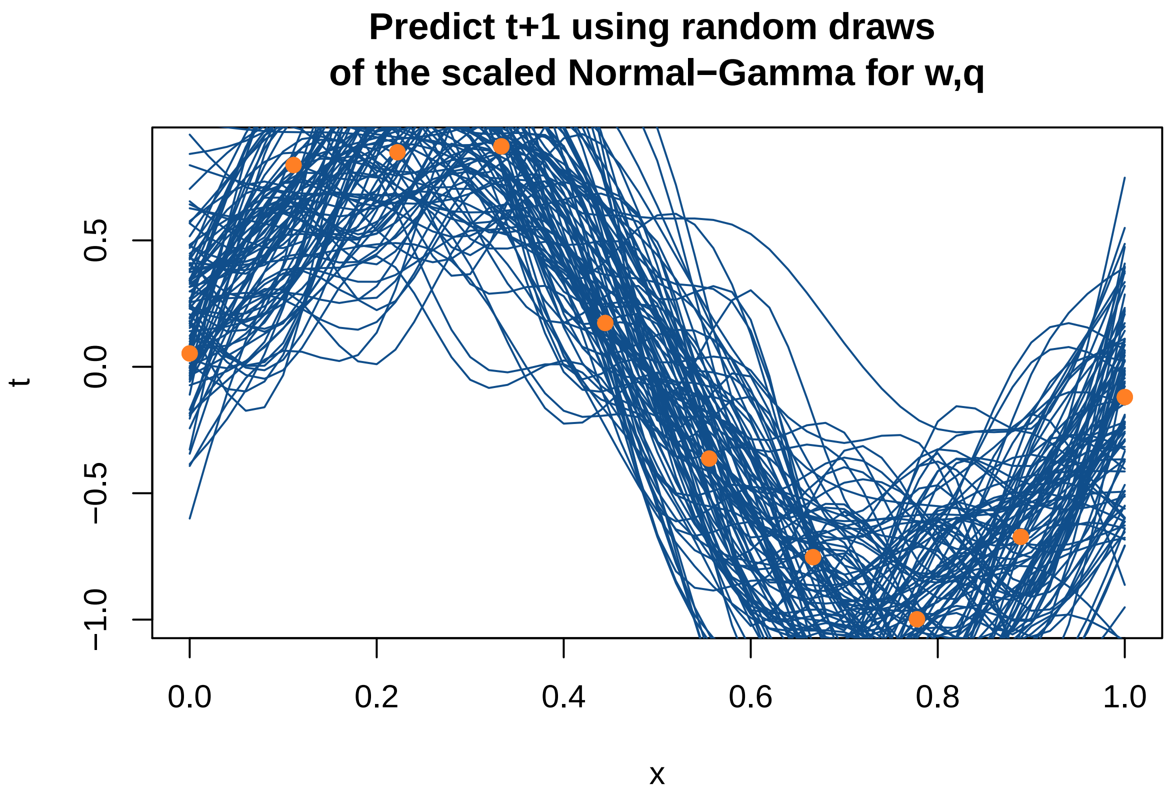
$a = -\frac{1}{2}, b = 0, \mu = 0, D = \delta I$

- Produce a figure that includes posterior draws of the linear predictor for this model

1. Continuation of your work on smooth function estimation with the data in curve_data.txt. Extend your program to also learn q from the data. Work with the Normal-Gamma prior with specification. Produce a figure that includes posterior draws of the linear predictor for this model


Model Options for Curve Data
mle
bayes
bayes (unknown precision)
c(−1, 1)
c(0, 1)
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0


Predict t+1 using student−T distribution
t
x
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0



Exploring the sensitvity to delta exposed that a large delta drives the values of the parameters to 0 and thus there is no correlation between x and the fitted values for t. A low delta or zero delta drives a closer fit to the observations.

2.1 Priors that penalize the L norm of the mean of a Gaussian and MAP estimation

1. Suppose for simplicity that a > 0 (a < 0 can be handled in the same way), assume also that $\lambda > 0$ and consider the function

$f(\mu) = (\mu - a)^2 + \lambda|\mu|$

Show that μ is minimised at $(a - \lambda/2)^+$ where x^+ denotes the positive part of x.

We take the derivative and equate to zero to find the λ that maximizes the L norm

$\dfrac{\partial f(\mu)}{\partial \mu} = 2(\mu - a)^2 + \lambda|\mu| = 0$

$\lambda = -2(\mu - a)$

$\mu = a - \dfrac{\lambda}{2}$

And we can check that it is indeed a minimum as the second derivative is > 0:

$\dfrac{\partial}{\partial \mu}(\dfrac{\partial f(\mu)}{\partial \mu}) = 2 > 0$

2. Find $\mathbf{w}_{MAP}$ under this model in closed form

$\mathbf{w}_{MAP} = arg\ max\ p(w|t) = arg\ max\ p(t|w) +\ p(w)$

We can take the log on both sides, as the monotonicity of the log function does not alter the solution for the maximixation,

$\mathbf{w}_{MAP} = arg\ max\ log\ p(w|t) = arg\ max\ log\ p(t|w) + log\ p(w)$

$$= \ arg\ max\ log\ \prod_n p(t_n|w) + log\ e^{-\frac{\delta}{2}\sum_i |w_i|}$$

$$= \ arg\ max\ \sum_n\ log\ N(w_n, q^{-1}I) + log\ e^{-\frac{\delta}{2}\sum_i |w_i|}$$

$$= \ arg\ max\ \sum_n\ const\ log\ e^{-\frac{1}{2}q(t_n-w)^T(t_n-w)} + log\ e^{-\frac{\delta}{2}\sum_i |w_i|}$$

$$= \ arg\ max\ \sum_n\ -\frac{const}{2}q(t_n-w)^T(t_n-w) - \frac{\delta}{2}\sum_i |w_i|$$

$$\propto \ min\ \sum_n\ q(t_n-w)^T(t-w) + \delta\sum_i |w_i|$$

$$= \ min\ q\sum_n\sum_i\ (t_n-w)^2 + \delta\sum_i |w_i|$$

Taking the derivative and equating to 0,

$\frac{\partial}{\partial w_i} = 0$

$$-2q\sum_n(t_{ni}-w_i) + \delta = 0$$

$$-Nw_i + \sum_n t_{ni} = \frac{\delta}{2}q^{-1}$$

Thus, the mode of the posterior distribution is at:

$$w_i = \frac{1}{N}(\sum_n t_{ni} - \frac{\delta}{2}q^{-1})$$