Exercise 1. Regression.

Q1. According to Theorem 4.4.1 Bayes rule for linear Gaussian System

$p(x|Y) = \mathcal{N}(x|\mu_{x|y}, \Sigma_{x|y})$

$\Sigma_{x|y}^{-1} = \Sigma_x^{-1} + A^T \Sigma_y^{-1} \cdot A$ ①

A is matrix of size $R^{d \times N}$
$p(y|x) = \mathcal{N}(y|Ax+b, \Sigma_y)$

$\mu_{x|y} = \Sigma_{x|y} \cdot [A^T \Sigma_y^{-1}(y-b) + \Sigma_x^{-1}\mu_x]$ ②

$\Sigma_Y^{-1} = \beta I$
$\mu_Y = V$
$\Sigma_x^{-1} = \alpha I$
$\mu_x = 0$

a)

$\because p(w|\vec{X}, \vec{t}, \beta, \alpha) \propto \underbrace{p(t|X, w, \beta)}_{Y} \cdot \underbrace{p(w|\alpha)}_{X}$

according to formula ①:

$\Sigma_N^{-1} = \Sigma_{w|x,t,\beta,\alpha}^{-1} = \Sigma_x^{-1} + A^T \cdot \Sigma_y^{-1} \cdot A = \alpha I + \Phi^T \cdot \beta \cdot \Phi$

according to formula ②:

$\mu_N = \mu_{w|x,t,\beta,\alpha} = \Sigma_N \cdot [A^T \cdot \Sigma_y^{-1} \cdot (y-b) + \Sigma_x^{-1}\mu_x]$

$= \Sigma_N \cdot [\Phi^T \cdot \beta \cdot \vec{t} + \alpha \cdot 0]$

$= \beta \Sigma_N \Phi^T \vec{t}$

b) Now the posterior is regarded as prior when adding data points (X_{N+1}, t_{N+1})

$p(w'|\mathbf{x}', \mathbf{t}', \beta, \alpha) \propto \underbrace{p(t_{N+1}|X_{N+1}, w, \beta)}_{Y} \cdot \underbrace{p(w|X, t, \beta, \alpha)}_{X}$

according to formula ①:

$\Sigma_{N+1}^{-1} = \Sigma_{w'|x',t',\beta,\alpha}^{-1} = \alpha I + \underline{\vec{\Phi}^T \beta \vec{\Phi} + \phi(x_{N+1}) \beta \phi(x_{N+1})}$

$= \alpha I + \beta \vec{\Phi}_{N+1}^T \vec{\Phi}_{N+1}$

according to formula ②

$\mu_{N+1} = \mu_{w'|x',t',\beta,\alpha} = \Sigma_{N+1} \cdot [A^T \cdot \Sigma_y^{-1} \cdot (y-b) + \Sigma_x^{-1}\mu_x]$

$= \Sigma_{N+1} \cdot [\phi^T(x_{N+1}) \beta \cdot t(x_{N+1}) + \Sigma_N^{-1} \cdot \beta \cdot \Sigma_N \Phi^T \vec{t}]$

$= \Sigma_{N+1} \cdot [\beta \cdot \phi_{N+1}^T t_{N+1} + \beta \vec{\Phi}^T \vec{t}]$

$= \Sigma_{N+1} \cdot \beta \cdot \vec{\Phi}_{N+1}^T t_{N+1}$

Q2. a). when $q=1$. $L_{q=1}(t, y(x)) = |y(x) - t|$

$E(L_1) = \int_{R^d}\int |y(x)-t| \cdot p(x,t) \cdot dt \cdot dx$

$= \int_{R^d} \underbrace{\int |y(x)-t| \cdot p(t|x) \cdot dt}_{y(x)} \cdot p(x) \cdot dx.$

When $y(x)$ is minimize, $E(L_1)$ will be minimal.

That is: $y(x) = \underset{\hat{y}}{\operatorname{argmin}} \int_R |\hat{y} - t| \, p(t|x) \, dt.$

the best y that can minimize the $y(x)$ is median of whole $\vec{t}$

b) for each point X, E is invariant with X,

so $\min E[L_q]$

$\Leftrightarrow \min \int_R |\hat{y}-t|^0 p(t|X)\,dt$

if $y \neq t$, then $|y-t|^0 = 1$,

$\Rightarrow \min \int_R p(t|X)\,dt$

if $y=t$, $|y-t|^0 = 0^0$ is a singularity

at this point, we obtain a big reduction.

so we just choose the point t with biggest frequency, namely $\underset{t \in R}{\arg\max}\, p(t|X)$, which get the biggest reduction.