1) a) (1) Arbitrary Gaussian:

$$\mathcal{N}(w|\mu_N,\Sigma_N) = \left((2\pi)^d \cdot |\Sigma_N|\right)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\cdot(w-\mu_N)^T \Sigma_N^{-1} \cdot (w-\mu_N)\right)$$

$$\Rightarrow \text{exp. term} = -\frac{1}{2}\cdot(w-\mu_N)^T \cdot \Sigma_N^{-1} \cdot (w-\mu_N)$$

$$= \underbrace{-\frac{1}{2}\cdot w^T \cdot \Sigma_N^{-1} \cdot w}_{\text{"quadratic term in } w\text{"}} \underbrace{- \mu_N^T \cdot \Sigma_N^{-1} \cdot w}_{\text{" linear term in } w\text{"}} + \text{const}$$

(2) Posterior Distribution:

- $p(t|X,w,\beta) = \prod_{n=1}^{N} \mathcal{N}(t_n | w^T \cdot \phi(x_n), \beta^{-1})$ (Likelihood)
- $p(w|\alpha) = \mathcal{N}(w|0, \alpha^{-1} \cdot I)$ (Prior)

$$\Rightarrow \underset{\text{(of posterior)}}{\text{exp. term}} = \sum_{n=1}^{N} -\frac{1}{2}\cdot(t_n - w^T\cdot\phi(x_n))^T \cdot \beta \cdot (t_n - w^T \cdot \phi(x_n))$$

$$-\frac{1}{2}\cdot(w-0)^T \cdot \alpha \cdot I \cdot (w-0)$$

$$= -\frac{\beta}{2} \cdot \sum_{n=1}^{N} (t_n - w^T \cdot \phi(x_n))^2 - \frac{\alpha}{2}\cdot w^T \cdot w$$

$$= -\frac{\beta}{2} \cdot \sum_{n=1}^{N} \left(t_n^2 - 2 t_n w^T \cdot \phi(x_n) + w^T \cdot \phi(x_n) \cdot \phi(x_n)^T w\right) - \frac{\alpha}{2} w^T w$$

$$= -\frac{1}{2} w^T \cdot \left(\sum_{n=1}^{N} \beta \cdot \phi(x_n) \cdot \phi(x_n)^T + \alpha \cdot I\right) \cdot w$$

$$- \sum_{n=1}^{N} \beta \cdot t_n \cdot \phi(x_n)^T \cdot w \quad + \text{const}$$

Compare quadratic and linear terms of (1) and (2):

- $\Sigma_N^{-1} \overset{(!)}{=} \sum_{n=1}^{N} \beta \cdot \phi(x_n) \cdot \phi(x_n)^T + \alpha \cdot I = \beta \cdot \Phi^T \cdot \Phi + \alpha \cdot I$
- $\mu_N^T \cdot \Sigma_N^{-1} \overset{(!)}{=} \sum_{n=1}^{N} \beta \cdot t_n \cdot \phi(x_n)^T = \beta \cdot t^T \cdot \Phi$

$$\Leftrightarrow \mu_N = \beta \cdot \Sigma_N \cdot \Phi^T \cdot t$$

(e) $\cdot\ p(t_{N+1} | x_{N+1}, w, \beta) = \mathcal{N}(t_{N+1} | w^T \cdot \phi(x_{N+1}), \beta^{-1})$ (Likelihood)

$\cdot\ p(w | X, t, \beta, \alpha) = \mathcal{N}(w | \mu_N, \Sigma_N)$ (Prior)

$\Rightarrow$ exp. term (of posterior) $= -\frac{1}{2}\beta \cdot (t_{N+1} - w^T \cdot \phi(x_{N+1}))^2 - \frac{1}{2}\cdot(w-\mu_N)^T \cdot \Sigma_N^{-1} \cdot (w - \mu_N)$

$$= -\frac{1}{2}\cdot w^T \cdot \underbrace{\left[\Sigma_N^{-1} + \beta \cdot \phi(x_{N+1}) \cdot \phi(x_{N+1})^T\right]}_{=\Sigma_{N+1}^{-1}} w$$

$$- \underbrace{\left[\beta \cdot \phi(x_{N+1}) \cdot t_{N+1} + \mu_N^T \cdot \Sigma_N^{-1}\right]}_{\overset{(!)}{=} \mu_{N+1}^T \cdot \Sigma_{N+1}^{T}} \cdot w + \text{const}$$

$$\Rightarrow \mu_{N+1} = \beta \cdot \Sigma_{N+1}^T \cdot \Phi^T t$$

2) $\mathbb{E}[L_q] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} |y(x) - t|^q \cdot p(x,t)\, dt\, dx$

$$= \int_{\mathbb{R}^d} p(x) \cdot \left[\int_{\mathbb{R}} |y(x) - t|^q \cdot p(t|x)\, dt \right] dx$$

$\Rightarrow$ For every $x \in \mathbb{R}^d$, choose $\hat{y} := y(x) \in \mathbb{R}$, such that

$$F_q(\hat{y}|x) := \int_{\mathbb{R}} |\hat{y} - t|^q \cdot p(t|x)\, dt \longrightarrow \min$$

(a) For $q = 1$, this gives

$$y(x) := \underset{\hat{y} \in \mathbb{R}}{\operatorname{argmin}} \left(F_q(\hat{y}|x) \right)$$

(b) <u>Assumption:</u> $p(\cdot|x)$ is uniformly continuous:

- Fix $0 < \varepsilon \ll 1$:

$\Rightarrow \exists\, 0 < \delta = \delta(\varepsilon) \ll 1 : \forall \hat{y} \in \mathbb{R},\ t \in U_\delta(0)$:

$$|p(\hat{y} + t|x) - p(\hat{y}|x)| < \varepsilon$$

- Consider

$$F_q(\hat{y}|x) = \int_{\mathbb{R}} |\hat{y} - t|^q \cdot p(t|x)\, dt = \int_{\mathbb{R}} |t|^q \cdot p(\hat{y} + t|x)\, dt$$

$$= \underbrace{\int_{\mathbb{R} \setminus U_\delta(0)} |t|^q \cdot p(\hat{y} + t|x)\, dt}_{\downarrow} + \int_{U_\delta(0)} |t|^q \cdot p(\hat{y} + t|x)\, dt$$

Choose $0 < q = q(\delta) \ll 1$:

$$\left| \int_{\mathbb{R} \setminus U_\delta(0)} |t|^q \cdot p(\hat{y} + t|x)\, dt - \int_{\mathbb{R} \setminus U_\delta(0)} p(\hat{y} + t|x)\, dt \right| < \varepsilon$$

$$\Rightarrow \int_{\mathbb{R} \setminus U_\delta(0)} |t|^q \cdot p(\hat{y} + t|x)\, dt = \int_{\mathbb{R} \setminus U_\delta(0)} p(\hat{y} + t|x)\, dt + O(\varepsilon)$$

$$= \int_{\mathbb{R}\setminus \mathcal{U}_\delta(0)} p(\hat{y}+t|x)\,dt + \int_{\mathcal{U}_\delta(0)} |t|^q \cdot p(\hat{y}+t|x)\,dt + \mathcal{O}(\varepsilon)$$

$$= \underbrace{\int_{\mathbb{R}} p(\hat{y}+t|x)\,dt}_{=1} - \underbrace{\int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot p(\hat{y}+t|x)\,dt} + \mathcal{O}(\varepsilon)$$

$$\left| \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot p(\hat{y}+t|x)\,dt - \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot p(\hat{y}|x)\,dt \right|$$

$$= \left| \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot \left[p(\hat{y}+t|x) - p(\hat{y}|x) \right] dt \right|$$

$$\leq \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot \underbrace{\left| p(\hat{y}+t|x) - p(\hat{y}+x) \right|}_{< \varepsilon \text{ by uniform continuity}} dt$$

$$< \varepsilon \cdot \int_{\mathcal{U}_\delta(0)} 1-|t|^q\,dt = \mathcal{O}(\varepsilon)$$

$$\Rightarrow \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot p(\hat{y}+t|x)\,dt = \int_{\mathcal{U}_\delta(0)} (1-|t|^q)\cdot p(\hat{y}|x)\,dt + \mathcal{O}(\varepsilon)$$

$$= \left(\int_{\mathcal{U}_\delta(0)} 1-|t|^q\,dt \right)\cdot p(\hat{y}|x) + \mathcal{O}(\varepsilon)$$

$$= 1 - \underbrace{\left(\int_{\mathcal{U}_\delta(0)} 1-|t|^q\,dt \right)}_{>0}\cdot p(\hat{y}|x) + \mathcal{O}(\varepsilon)$$

$$\Rightarrow \underset{\hat{y}\in\mathbb{R}}{\operatorname{arg\,min}} \left(F_q(\hat{y}|x) \right) = \underset{\hat{y}\in\mathbb{R}}{\operatorname{arg\,max}} \left(p(\hat{y}|x) \right) =: y(x)$$