1.1 $$L(w) = \prod_{n=1}^{N} p(t_n | x_n, w)$$

$$= \prod_{n=1}^{N} \left[\sqrt{\frac{\beta}{2\pi}} \, e^{-\frac{\beta}{2}\left(t_n - \frac{w}{x_n}\right)^2} \right]$$

1.2 Define $E(w) = -\log L(w)$

$$= -\sum_{n=1}^{N} \left[\frac{1}{2}\log\beta - \frac{1}{2}\log(2\pi) - \frac{\beta}{2}\left(t_n - \frac{w}{x_n}\right)^2 \right]$$

$$= -\frac{N}{2}\log\beta + \frac{N}{2}\log(2\pi) + \frac{\beta}{2}\sum_n \left(t_n - \frac{w}{x_n}\right)^2$$

1.3 $$\frac{\partial E(w)}{\partial w} = \frac{\beta}{\cancel{2}} \sum_n \left[\cancel{2}\left(t_n - \frac{w}{x_n}\right)\left(\frac{-1}{x_n}\right) \right]$$

$$= \beta \sum_n \left[-\frac{t_n}{x_n} + \frac{w}{x_n^2} \right]$$

Setting to 0 for $\hat{w}_{ML}$:

$$\cancel{\beta} \sum_n \left[-\frac{t_n}{x_n} + \frac{\hat{w}_{ML}}{x_n^2} \right] = 0$$

$$\sum_n \frac{\hat{w}_{ML}}{x_n^2} = \sum_n \frac{t_n}{x_n}$$

$$\Rightarrow \boxed{\hat{w}_{ML} = \frac{\sum_n (t_n/x_n)}{\sum_n (1/x_n^2)}}$$

1.4 For single data pt., $\hat{w}_{ML} = \frac{t/x}{1/x^2} = tx$

So it is the product of distance between the districts and the connection density between them, which is the natural estimate of w

For multiple data pts., numerator and denominator both sum over the contributions from individual data pts. But it doesn't take the simpler intuitive form of $\dfrac{\sum_n t_n}{\sum_n (1/x_n)}$ because of the squared error, which makes $\hat{w}_{ML}$ more sensitive to pts. with smaller values of x (since in these cases, error is more sensitive to small deviations in w).

1.5 $p(w \mid X, \underline{t}) \propto p(\underline{t} \mid X, w)\, p(w)$

(for the discriminative modelling scenario)

$$\propto L(w)\, p(w)$$

$$\propto \left[\prod_{n=1}^{N} \sqrt{\frac{\beta}{2\pi}}\, e^{-\frac{\beta}{2}\left(t_n - \frac{w}{x_n}\right)^2} \right] \left[\sqrt{\frac{\alpha}{2\pi}}\, e^{-\frac{\alpha}{2} w^2} \right]$$

1.6 Define $\tilde{E}(w) = -\log p(w \mid X, \underline{t})$

$$= -\frac{N}{2}\log\beta + \frac{N}{2}\log(2\pi) + \frac{\beta}{2}\sum_n \left(t_n - \frac{w}{x_n}\right)^2$$

$$\cancel{+} -\frac{1}{2}\log\alpha + \frac{1}{2}\log(2\pi) + \frac{\alpha}{2}w^2 + \underbrace{K}_{\text{(norm. const.)}}$$

(Combining all constants) $= \dfrac{\beta}{2}\sum_n \left(t_n - \dfrac{w}{x_n}\right)^2 + \dfrac{\alpha}{2} w^2 + K'$

1.7 $$\frac{\partial \tilde{E}(w)}{\partial w} = \frac{\beta}{\cancel{2}} \sum_n \left[\cancel{2}\left(t_n - \frac{w}{x_n}\right)\left(-\frac{1}{x_n}\right)\right] + \frac{\alpha}{\cancel{2}}(\cancel{2}w)$$

$$= \beta \sum_n \left[\frac{-t_n}{x_n} + \frac{w}{x_n^2} \right] + \alpha w$$

Setting to 0 for $\hat{w}_{MAP}$:

$$\beta \sum_n \left[\frac{-t_n}{x_n} + \frac{\hat{w}_{MAP}}{x_n^2} \right] + \alpha \hat{w}_{MAP} = 0$$

$$\beta \sum_n \frac{\hat{w}_{MAP}}{x_n^2} + \alpha \hat{w}_{MAP} = \beta \sum_n \frac{t_n}{x_n}$$

$$\hat{w}_{MAP} \left[\beta \sum_n \frac{1}{x_n^2} + \alpha \right] = \beta \sum_n \frac{t_n}{x_n}$$

$$\boxed{\hat{w}_{MAP} = \frac{\sum_n (t_n/x_n)}{\sum_n (1/x_n^2) + \alpha/\beta}}$$

1.8 Prior is favoring smaller (magnitude) values of w; its strength is controlled by α (larger α = stronger prior).

2.1 $\nabla_{\underline{w}} \tilde{E}_n(\underline{w})$ to be found, where

$$\tilde{E}_n(\underline{w}) = \frac{1}{2}(t_n - \underbrace{y(x_n; \underline{w})}_{\triangleq y_n})^2 + \frac{\lambda}{2} \|\underline{w}\|^2$$

$$= \frac{1}{2}(t_n - y_n)^2 + \frac{\lambda}{2}(w_0^2 + w_1^2)$$

So, $$\frac{\partial \tilde{E}_n(\underline{w})}{\partial w_0} = \frac{1}{\cancel{2}}(-\cancel{2})(t_n - y_n)\cdot\frac{\partial y_n}{\partial w_0} + \frac{\lambda}{\cancel{2}}(\cancel{2}w_0)$$

$$= (y_n - t_n)(1) + \lambda w_0$$

$$\frac{\partial \tilde{E}_n(\underline{w})}{\partial w_1} = \frac{1}{\cancel{2}}(-\cancel{2})(t_n - y_n)\frac{\partial y_n}{\partial w_1} + \frac{\lambda}{\cancel{2}}(\cancel{2}w_1)$$

$$= (y_n - t_n)(x_n + 2w_1 x_n^2) + \lambda w_1$$

$$\nabla_{\underline{w}} \tilde{E}_n(\underline{w}) = \begin{pmatrix} y_n - t_n + \lambda w_0 \\ (y_n - t_n)(x_n + 2w_1 x_n^2) + \lambda w_1 \end{pmatrix}$$

2.2 Using n to denote the index of the data pt. chosen at iteration $\tau+1$:

$$w_0^{(\tau+1)} = w_0^{(\tau)} - \eta \left.\frac{\partial \tilde{E}_n(\underline{w})}{\partial w_0}\right|_{\underline{w}^{(\tau)}}$$

$$= w_0^{(\tau)} - \eta\left(y_n - t_n + \lambda w_0^{(\tau)}\right)$$

$$w_1^{(\tau+1)} = w_1^{(\tau)} - \eta \left.\frac{\partial \tilde{E}_n(\underline{w})}{\partial w_1}\right|_{\underline{w}^{(\tau)}}$$

$$= w_1^{(\tau)} - \eta\left[(y_n - t_n)(x_n + 2w_1^{(\tau)} x_n^2) + \lambda w_1^{(\tau)}\right]$$

Here $y_n \triangleq y(x_n; \underline{w}^{(\tau)})$

3. (a), (d), (e)