1)

a) $$\sigma'(x) = \frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right) = e^{-x}(1+e^{-x})^{-2} = \frac{e^{-x}}{(1+e^{-x})^2}$$

Same

$$\sigma(x)(1-\sigma(x)) = \frac{1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2} = \frac{e^{-x}+1-1}{(1+e^{-x})^2}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

b) $$nll(\theta) = -\sum_i y_i \log\sigma(\theta^T x_i) + (1-y_i)\log(1-\sigma(\theta^T x_i))$$

$$\nabla_\theta nll(\theta) = -\sum_i y_i \frac{1}{\sigma(\theta^T x_i)}\sigma'(\theta^T x_i) + (1-y_i)\frac{1}{1-\sigma(\theta^T x_i)}(-\sigma'(\theta^T x_i))$$

$$= -\sum_i y_i(1-\sigma(\theta^T x_i))x_i - (1-y_i)\sigma(\theta^T x_i)x_i$$

$$= -\sum_i y_i x_i - y_i\sigma(\theta^T x_i)x_i - \sigma(\theta^T x_i)x_i + y_i\sigma(\theta^T x_i)x_i$$

$$= \sum_i (\sigma(\theta^T x_i) - y_i)x_i$$

Let $u_i = \sigma(\theta^T x_i)$

Then

$$\sum_i (u_i - y_i)x_i = (u-y)X = X^T(u-y)$$

c) $H = X^T S X$

$$H_\theta = \nabla_\theta (\nabla_\theta \, nll(\theta))^T = \nabla_\theta [X^T(u-y)]^T$$

$$= \nabla_\theta (u^T X - y^T X) = \nabla_\theta u^T X - \nabla_\theta y^T X$$

$$\nabla_\theta y^T X \to 0$$

$$\nabla_\theta u^T X \to \nabla_\theta \sum_i \sigma(\theta^T x_i) X$$

$$= \nabla_\theta \sigma(X\theta)^T X$$

use results from part a

$$= X^T \sigma(1-\sigma) X$$

$$\text{diag } u(1-u) = X^T S X$$

now just show this is semi definite positive.

Since $u_i = \sigma(\theta^T x_i)$

σ of anything is between 0 and 1

So $u(1-u)$ must be positive definite.

2.

$$p(x, \sigma^2) = \frac{1}{Z} e^{-\frac{x^2}{2\sigma^2}}$$

$$\int p(x, \sigma^2) = \int \frac{1}{Z} e^{-\frac{x^2}{2\sigma^2}} = 1$$

gaussian integral

$$= \frac{1}{Z} \sqrt{\frac{\pi}{\frac{1}{2\sigma^2}}} = \frac{1}{Z} \sqrt{\pi \cdot 2\sigma^2} = 1$$

$$Z = \sqrt{\pi \cdot 2}\, \sigma$$

3) since $\mathcal{N}(y_i \mid u_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - u_i)^2}{2\sigma^2}\right)$

$$\log \mathcal{N}(y_i \mid u_i, \sigma^2) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(y_i - u_i)^2}{2\sigma^2}\right)$$

$$= -\frac{1}{2\sigma^2}(y_i - u_i)^2 + \text{constant}$$

so

$$\log \mathcal{N}(y_i \mid w_0 + w^T x_i, \sigma^2) = -\frac{1}{2\sigma^2}(y_i - w_0 - w^T x_i)^2 + \text{constant}$$

$\Downarrow$

$$\log \mathcal{N}(w_j \mid 0, \tau^2) = -\frac{1}{2\tau^2} w_j^2 + \text{const}$$

$$\arg\max_w \left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - w_0 - w^T x_i)^2 - \frac{1}{2\tau^2}\sum_{j=1}^{D} w_j^2\right)$$

$$= \arg\min_w \left(\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - w_0 - w^T x_i)^2 + \frac{1}{2\tau^2}\sum_{j=1}^{D} w_j^2\right)$$

$$\frac{1}{2\tau^2}\sum_{j=1}^{D} w_j^2 = \lambda \|w\|_2^2$$

$\Downarrow$

$$\arg\min \sum_{i=1}^{N}(y_i - w_0 - w^T x_i)^2 + \lambda \|w\|_2^2$$

b)

$$||Ax - b||_2^2 + ||\Gamma x||_2^2$$

$$\nabla_x f = \nabla_x ||Ax-b||_2^2 - ||\Gamma x||_2^2$$

$$= \nabla_x \left| (Ax - b)^T (Ax - b) + (\Gamma x)^T (\Gamma x) \right|$$

$$= \nabla_x \left((A^T x^T - b^T)(Ax - b) + \Gamma^T x^T \Gamma x \right)$$

$$= \nabla_x \left(x^T A^T A x - 2x^T A^T b + b^T b + x^T \Gamma^T \Gamma x \right)$$

$$= 2A^T A x - 2A^T b + 2\Gamma^T \Gamma x = 0$$

$$A^T b = x(A^T A + \Gamma^T \Gamma)$$

$$x = A^T b (A^T A + \Gamma^T \Gamma)^{-1}$$

d) $\| Ax + b\mathbb{1} - y\|_2^2 + \|\Gamma x\|_2^2$

$$= (Ax + b\mathbb{1} - y)^T (Ax + b\mathbb{1} - y) + (\Gamma x)^T(\Gamma x)$$

$$(x^T A^T + b\mathbb{1}^T - y^T)(Ax + b\mathbb{1} - y) + x^T\Gamma^T\Gamma x$$

$$x^T A^T Ax + 2b\mathbb{1}^T Ax - 2y^T Ax - 2b\mathbb{1}^T y + b^2 n + y^T y + x^T\Gamma^T\Gamma x$$

$$\nabla_x f = 2A^T Ax + 2bA^T\mathbb{1} - 2A^T y + 2\Gamma^T\Gamma x = 0$$

$$\nabla_b f = 2\mathbb{1}^T Ax - 2\mathbb{1}^T y + 2bn = 0$$

$$b = \frac{\mathbb{1}^T(y - Ax)}{n} \quad \rightarrow \text{plug that back in}$$

$$(A^T A + \Gamma^T\Gamma)x + \left(\frac{\mathbb{1}^T(y - Ax)}{n}\right)A^T\mathbb{1} - A^T y = 0$$

$$(A^T A + \Gamma^T\Gamma)x + \frac{1}{n}A^T\mathbb{1}\mathbb{1}^T y - \frac{1}{n}A\mathbb{1}\mathbb{1}^T Ax - A^T y = 0$$

$$x\left(A^T A + \Gamma^T\Gamma - \frac{1}{n}A^T\mathbb{1}\mathbb{1}^T A\right) = A^T y - \frac{1}{n}A^T\mathbb{1}\mathbb{1}^T y$$

$$x = \frac{A^T y - \frac{1}{n}A^T\mathbb{1}\mathbb{1}^T y}{A^T A + \Gamma^T\Gamma - \frac{1}{n}A^T\mathbb{1}\mathbb{1}^T A}$$