(1) Note that $y_i \in \{0,1\}$ $\forall i$,

$$\Rightarrow \sum_{i=1}^{n} \mathbb{I}(y_i=1)\log\sigma(w^T x_i) + \mathbb{I}(y_i=0)\log(1-\sigma(w^T x_i))$$

$$= \sum_{i=1}^{n} y_i \log\sigma(w^T x_i) + (1-y_i)\log(1-\sigma(w^T x_i))$$

Realize $\sigma(w^T x_i) = \dfrac{1}{1+\exp(-w^T x_i)}$

$$\Rightarrow \frac{\partial}{\partial w}\sigma(w^T x_i) = \frac{x_i \exp(-w^T x_i)}{(1+\exp(-w^T x_i))^2}$$

$$= \sigma(w^T x_i)(1-\sigma(w^T x_i))x_i$$

$$\Rightarrow \frac{\partial}{\partial w}\log\sigma(w^T x_i) = \frac{1}{\sigma(w^T x_i)}\frac{\partial}{\partial w}\sigma(w^T x_i)$$

$$= (1-\sigma(w^T x_i))x_i$$

Similarly,

$$\frac{\partial}{\partial w}\log(1-\sigma(w^T x_i)) = \frac{1}{1-\sigma(w^T x_i)}\frac{\partial}{\partial w} -\sigma(w^T x_i)$$

$$= \frac{1}{1-\sigma(w^T x_i)}\left(-\sigma(w^T x_i)(1-\sigma(w^T x_i))x_i\right)$$

$$= -\sigma(w^T x_i)x_i$$

$$\cancel{\Rightarrow \nabla f(w) = \sum_{i=1}^{n} y_i x_i (1-\sigma(w^T x_i))}$$

$$\Rightarrow \nabla f(w) = \sum_{i=1}^{n} y_i x_i (1 - \sigma(w^T x_i)) - (1 - y_i) x_i \sigma(w^T x_i)$$

$$= \sum_{i=1}^{n} x_i \Big(y_i \Pr_w(Y = 0 \mid X = x_i) - (1 - y_i) \Pr_w(Y = 1 \mid X = x_i) \Big)$$

This means that for each x_i,

if $y_i = 1$, then $\nabla f(w) \big|_{X = x_i} = x_i \Pr_w(Y = 0 \mid X = x_i)$

if $y_i = 0$, then $\nabla f(w) \big|_{X = x_i} = - x_i \Pr_w(Y = 1 \mid X = x_i)$

② $\lambda_i + 2\lambda, \quad \forall i \in [d].$

③. $\max_w \log p(w) + \log p(D|w)$

$$\iff \max_w -\frac{1}{b}\|w\|_1 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - w^T x_i)^2$$

$$\iff \min_w \frac{1}{2}\sum_{i=1}^{n}(y_i - w^T x_i)^2 + \frac{\sigma^2}{b}\|w\|_1$$

$$\iff \min_w \frac{1}{2}\|Xw - y\|_2^2 + \frac{\lambda}{\cancel{2}}\|w\|_1,$$

where $\lambda := \sigma^2 / b$.

$\blacksquare$.