1.1

(a)

$$p(D|w) = \prod_{i=1}^{n} p(y_i|x_i, w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \cancel{w^T x_i}\ x_i^T w)^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(y_i - \cancel{w^T x_i}\ x_i^T w)^2}{2\sigma^2}\right\}$$

likelihood

$$l(w) = \log P(D|w) = n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{\sum_{i=1}^{n}(y_i - w^T x_i)^2}{2\sigma^2}$$

(b)

$$\frac{\partial l(w)}{\partial w} = -\frac{2}{2\sigma^2}\sum_{i=1}^{n}(-x_i)(y_i - x_i^T w) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n}(x_i x_i^T) w = -\frac{1}{\sigma^2}(X^T y - X^T X w) = 0$$

$$\Rightarrow \hat{w}_{MLE} = (X^T X)^{-1} \cdot X^T y$$

(c)

$$\underset{\hat{w}}{\arg\max}\ l(w) = \underset{\hat{w}}{\arg\max}\ -n \cdot \log\sqrt{2\sigma^2\pi} + \sum_{i=1}^{n}\left[\frac{-1}{2\sigma^2}(y_i - x_i^T w)^2\right]$$

$$\Rightarrow \underset{\hat{w}}{\arg\max} \sum_{i=1}^{n} \frac{-1}{2\sigma^2}(y_i - x_i^T w)^2 \Rightarrow \underset{\hat{w}}{\arg\max} \sum_{i=1}^{n} -\frac{1}{2}(y_i - x_i^T w)^2 \Rightarrow \underset{\hat{w}}{\arg\max} \sum_{i=1}^{n} -(y_i - x_i^T w)^2$$

equals to

$$\underset{\hat{w}}{\arg\min}\ L(w) = \underset{\hat{w}}{\arg\min}\ \frac{1}{n}\|X \cdot w - y\|_2^2 = \underset{\hat{w}}{\arg\min}\ \|X \cdot w - y\|_2^2$$

(d)

$$\nabla_w L(w) = \frac{2}{n}\|X \cdot w - y\| \cdot X$$

$$w = w - \eta \nabla_w L(w) = w - \eta \cdot \left(\frac{2}{n}\|X \cdot w - y\| \cdot X\right)$$

(e)

$$w = w - \gamma^{(t)} \nabla J_i(\theta) = w - \gamma \cdot (-y^{(i)} - \theta^T x^{(i)})\ x_j^{(i)}$$