2) log-likelihood gradient and Hessian

a) $f(x) = \phi(x)^T \beta$, $p(x) = \sigma(f(x))$, $\sigma(z) = 1/(1+e^{-z})$

$$L(\beta) = -\sum_{i=1}^{n} \left[y_i \log p(x) + (1-y_i) \log(1-p(x_i)) \right]$$

$$= -\sum_{i=1}^{n} \left[y_i \cdot \log(\sigma(\phi(x_i)^T\beta)) + (1-y_i) \log(1-\sigma(\phi(x_i)^T\beta)) \right]$$

$$\frac{\partial}{\partial \beta} L(\beta) = -\sum_{i=1}^{n} \left[y_i \cdot \frac{1}{\sigma(\phi(x_i)^T\beta)} \cdot \sigma(\phi(x_i)^T\beta)(1-\sigma(\phi(x_i)^T\beta)) \cdot \phi(x_i)^T \right.$$
$$- (1-y_i) \cdot \frac{1}{1-\sigma(\phi(x_i)^T\beta)} \cdot \sigma(\phi(x_i)^T\beta)(1-\sigma(\phi(x_i)^T\beta))$$
$$\left. \cdot\, \phi(x_i)^T \right]$$

$$= -\sum_{i=1}^{n} \left[y_i \cdot (1-\sigma(\phi(x_i)^T\beta)) \cdot \phi(x_i)^T \right.$$
$$\left. - (1-y_i) \cdot \sigma(\phi(x_i)^T\beta) \cdot \phi(x_i)^T \right]$$

$$= -\sum_{i=1}^{n} \left[(y_i - y_i \cdot \sigma(\phi(x_i)^T\beta)) \cdot \phi(x_i)^T \right.$$
$$\left. - (\sigma(\phi(x_i)^T\beta) - y_i \sigma(\phi(x_i)^T\beta)) \cdot \phi(x_i)^T \right]$$

$$= -\sum_{i=1}^{n} \phi(x_i)^T \left[y_i - y_i \sigma(\phi(x_i)^T\beta) - \sigma(\phi(x_i)^T\beta) \right.$$
$$\left. + y_i \sigma(\phi(x_i)^T\beta) \right]$$

$$= -\sum_{i=1}^{n} \phi(x_i)^T \left[y_i - \sigma(\phi(x_i)^T\beta) \right]$$

$$= \sum_{i=1}^{n} \left[\sigma(\phi(x_i)^T\beta) - y_i \right] \cdot \phi(x_i)^T$$

$$= \Phi^T \cdot (p - y)$$

b) $\frac{\partial^2}{\partial \beta^2} L(\beta) = \frac{\partial}{\partial \beta} \quad \frac{\partial}{\partial \beta} L(\beta)^T$

$$= \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \left((p(x_i) - y_i) \cdot \phi(x_i)^T \right)^T$$

$$= \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \phi(x_i) \cdot (p(x_i) - y_i)^T$$

$$= \sum_{i=1}^{n} \phi(x_i) \cdot \left(\sigma(\phi(x_i)^T \beta)(1 - \sigma(\phi(x_i)^T \beta)) \cdot \phi(x_i)^T \right)^T$$

$$= \Phi \cdot \left(p \cdot (1-p) \cdot \Phi^T \right)^T$$