1.8

a) $Y \mid X \sim N(X\beta, \sigma^2 I)$

i. $L(\beta) = \log p(Y \mid X, \beta, \sigma^2 I)$

$$= -\frac{1}{2}\log 2\pi |\sigma^2 I| - \frac{1}{2}(Y - X\beta)^T(\sigma^2 I)^{-1}(Y - X\beta)$$

ii. $\nabla_\beta L(\beta) = 0 \Rightarrow 0 = -\frac{1}{2}X^T(Y - X\beta)$

$$= X^T Y - X^T X \beta$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

The ML estimate is the same as the least-square estimate.

This shows minimizing MSE $\Leftrightarrow$ maximum likelihood w/ iid Gaussian noise