$$p(\theta) = \mathcal{N}(\theta;\, \mu_0, S_0) \qquad p(y|\theta) = \mathcal{N}(y;\, \phi(x)\theta,\, \sigma^2 I)$$

$y \in \mathbb{R}^N$
$X \in \mathbb{R}^{N\times D}$
$\phi(x) \in \mathbb{R}^{N\times M}$
$\theta \in \mathbb{R}^M$

Method 1: Differentiation

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \log p(y|\theta)p(\theta) = \underset{\theta}{\operatorname{argmax}} \log p(y|\theta) + \log p(\theta)$$

$$= \operatorname{argmax} -\tfrac{1}{2}\log 2\pi|\sigma^2 I| - \tfrac{1}{2\sigma^2}(y-\phi(x)\theta)^T(y-\phi(x)\theta) - \tfrac{1}{2}\log 2\pi|S_0| - \tfrac{1}{2}(\theta-\mu_0)^T S_0^{-1}(\,)$$

$$= \underset{\theta}{\operatorname{argmax}}\, L(\theta)$$

$$\frac{\partial}{\partial\theta} L(\theta) = -\tfrac{1}{\sigma^2}(y-\phi(x)\theta)^T \cdot -\phi(x) - (\theta-\mu_0)^T S_0^{-1} = 0$$

Fine for this to be $\frac{d}{d\theta}$ too in this case...

$$\tfrac{1}{\sigma^2} y^T\phi(x) - \tfrac{1}{\sigma^2}\theta^T\phi(x)^T\phi(x) - \theta^T S_0^{-1} + \mu_0^T S_0^{-1} = 0$$

$$\theta^T(\phi(x)^T\phi(x)\cdot\tfrac{1}{\sigma^2} + S_0^{-1}) = \tfrac{1}{\sigma^2} y^T\phi(x) + \mu_0^T S_0^{-1}$$

$$\hookrightarrow \theta^T = \left[\tfrac{1}{\sigma^2} y^T\phi(x) + \mu_0^T S_0^{-1}\right]\cdot\left[\phi(x)^T\phi(x)\cdot\tfrac{1}{\sigma^2} + S_0^{-1}\right]^{-1}$$

$$\theta^* = \left[\phi(x)^T\phi(x)\cdot\tfrac{1}{\sigma^2} + S_0^{-1}\right]^{-1}\left(S_0^{-1}\mu_0 + \tfrac{1}{\sigma^2}\phi(x)^T y\right)$$

Remember: You may need to be more explicit about the differentiation, if asked. Make sure you know the shapes!

Method 2: Find posterior.

$$y = \phi(x)\theta + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I_N)$$

↳ Implied by the likelihood density given, as discussed in lectures.

$$p\left(\begin{bmatrix}\theta\\ y\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\theta\\ y\end{bmatrix}; \begin{bmatrix}\mu_0\\ \phi(x)\mu_0\end{bmatrix}, \begin{bmatrix}S_0 & S_0\phi(x)^T\\ \phi(x)S_0 & \phi(x)S_0\phi(x)^T + \sigma^2 I\end{bmatrix}\right)$$

$$p(\theta|y) = \mathcal{N}\left(\theta;\, \mu_0 + S_0\phi(x)^T\left[\phi(x)S_0\phi(x)^T + \sigma^2 I\right]^{-1}(y-\phi(x)\mu_0),\; S_0 - \ldots\right)$$

↳ We don't need this.

In the Gaussian, the maximum is the mean, so...

$$\underset{\theta}{\operatorname{argmax}} \log p(\theta|y) = \underset{\theta}{\operatorname{argmax}}\, p(\theta|y) = \mu_0 + S_0\phi(x)^T[\ldots]^{-1}(y-\phi(x)\mu_0)$$

This is correct as is, but it can be shown to be equal to the solution above through the Woodbury identity. Added for completeness. Matrix cookbook eq 159:

$$S_0\phi(x)^T\left(\phi(x)S_0\phi(x)^T + \sigma^2 I\right)^{-1} = \left[S_0^{-1} + \sigma^{-2}\phi(x)^T\phi(x)\right]^{-1}\phi(x)^T\sigma^{-2}$$

$$\hookrightarrow \theta^* = \mu_0 + \left[S_0^{-1} + \sigma^{-2}\phi(x)^T\phi(x)\right]^{-1}\phi(x)^T\sigma^{-2}(y - \phi(x)\mu_0)$$

To get μ_0 to match we need another Woodbury...