Problem 1

	Larger C	Smaller C
Model capacity (large/small?)	large	small
Overfitting/Underfitting?	Overfitting	Underfitting
Bias variance (how/low?)	low bias / high variance	high bias / Low variance

Problem 2.

$$w \sim \mathcal{N}(0, \sigma_w^2)$$

$$P(w) = \frac{1}{\sqrt{2\pi}\,\sigma_w} \exp\left\{-\frac{1}{2}\left(\frac{w}{\sigma_w}\right)^2\right\}$$

$$P(w|D) = \frac{P(w) \cdot P(D|w)}{P(D)} \propto P(w) \cdot P(D|w)$$

$$P(w) = \prod_{i=0}^{d} P(w_i) \propto \prod_{i=0}^{d} \exp\left(-\frac{w_i^2}{2\sigma^2}\right) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=0}^{d} w_i^2\right\}$$

$$\hat{w}_{MAP} = \arg\max_w P(w)\, P(D|w)$$

$$= \arg\max_w \left[\log P(D|w) + \log P(w)\right]$$

$$= \arg\max_w \left[\underbrace{-\frac{1}{2\sigma_\epsilon^2}\sum_{m=1}^{M}(t - w^T x^{(m)})^2}_{\text{Likelihood + const}} \underbrace{- \frac{1}{2\sigma_w^2}\sum_{i=1}^{d} w_i^2}_{\text{Prior + const}} + \text{const}\right]$$

$$= \arg\min_w \left[\underbrace{\frac{1}{2M}\sum_{i=1}^{M}(t^{(i)} - w^T x^{(i)})^2}_{\text{MSE}} + \underbrace{\lambda\sum_{i=1}^{d} w_i^2}_{L_2\text{-penalty}}\right] \quad \text{for some } \lambda > 0$$

Prior

Posterior $\propto$ Likelihood $\times$ prior

$\Rightarrow$ The posterior of w is also a Gaussian distribution.

Problem 3.

$$\hat{w}_{MAP} = \underset{w}{\operatorname{argmax}}\ P(w)\,p(D|w)$$

$$= \underset{w}{\operatorname{argmax}}\ \log \frac{P(D|w)P(w)}{P(D)}$$

$$= \underset{w}{\operatorname{argmax}}\ \log P(D|w)P(w)$$

$$= \underset{w}{\operatorname{argmax}}\ \Big[\underbrace{\log P(D|w)}_{\text{log-likelihood}} + \underbrace{\log P(w)}_{\text{regularizer}}\Big]$$

By Laplace distribution,

$$f(x|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

$$= \frac{1}{2b}\begin{cases} \exp\left(-\frac{\mu-x}{b}\right) & \text{if } x<\mu \\ \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x\geq\mu \end{cases}$$

$$\text{log-likelihood} \Rightarrow -\sum_{d}\log P(w)$$

$$= -\sum_{d}\log\left(\frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)\right)$$

$$= \sum_{d}\frac{1}{2b}\left(\frac{|x-\mu|}{b}\right)$$