$\hat{\theta}_{OLS} = \underset{\theta}{\text{argmin}}\{(\Phi\theta - \mathcal{D})^T(\Phi\theta - \mathcal{D})\}$, which in both cases correspond to having $\lambda = 0$, or more precisely, either $\sigma^2 = 0$ or $\sigma_0 = \infty$. First, I'll discard $\sigma^2 = 0$. Sure enough this gives $\lambda = 0$ Instead, considering $\sigma_0 = \infty$ gives an infinitely wide prior. To me, a normal/Laplace distribution with infinite standard deviation (or variance) would be a uniform distribution on the interval $(-\infty, \infty)$. Using such a prior, i.e. $\mathcal{U}(-\infty, \infty)$, would effectively do nothing since all θ:s have an equal probability. Consequently, the posterior is only be affected by the likelihood. Referring to the lingo in the book, having the parameters $\sigma_0 = \infty$ and $\mu = 0$ corresponds to the non-informative prior, and hence the point estimate would be the same as OLS.

In conclusion, having an infinitely wide prior or a delta-function likelihood, the posterior will effecti

non-informative prior.

In the other case the likelihood variance $\sigma^2 = 0$, so the likelihood is then a delta function. The likelihood is a measure how well our model aligns with observed data.

If the likelihood is a delta function, then I guess the model aligns perfectly the data. The posterior will in this case also be a delta function.

Having a prior lets us add additional information about the data-generating process, we'd like to describe with our model. Using a non-informative prior, will