11.7.1 ~ 11.7.3 Priors and Posteriors
We know that MLE can be transformed into various forms of least square problem under different gaussian assumptions about the noise, and MAP can be transformed into a regularized version of the least squares problem. Here in the book, we assume that the noise is i.i.d. gaussian and the prior is also an i.i.d. gaussian vector; this would result in the posterior to be also gaussian random vectors. We see that for each bayesian update process, the variance of the posterior distribution is lowered and finally reaches a delta function. This makes sense since our model is learning from the data and trying to find the pattern or the real distribution of the training data.

11.7.4 Posterior Predictive
Most of the time, computing the posterior distribution is time-consuming so we may opt to compute the MAP estimator and plug in to the posterior predictive. Generally, as we have more data and done many rolls of bayesian updating processes, the variance of the posterior distribution will be small. If we have a weak prior (with a large variance), then the posterior distribution will be very close to the likelihood distribution of the data given the prior. If the prior is bad but strong (having a smaller variance), then the posterior predictive model may not be that effective.

11.7.5 Centering Data
It helps to reduce the computational difficulties in dealing with elongated gaussian contours. Converting from one coordinate to the other is easy and can be done in polynomial time.

11.7.6 Dealing with Multicollinearity
We also learn that the existence of multicollinearity could strongly affect the interpretation of the model parameter. In the book, we see that when we have such a problem, the posterior distribution of the parameters will be highly correlated with each other. This is important since feature selection needs to consider this.

11.7.7 ARD(Not fully understood)
This part is confusing without further examples, but this YouTube video https://www.youtube.com/watch?v=2gT-Q0NZzoE gives me some intuition about this method. Generally, it is trying to reconstruct the weights trying to make some of the weights zero even if we assume the weight distributions to be gaussian(which tends to push weights gradually to 0 at the same rate). I think in order to fully understand the mechanisms behind, phd level knowledge is required and as master we can just use sklearn package to perform this regression techniques.