In the wikipedia , the slides of Bayesian linear regression gives us the following formula:

IMG_256

In our model, we need to replace the square of error with the scale parameter alpha.

And the based on the formula of conditional probability, we can derive that

P(beta|alpha) \* P(alpha) = P(beta,alpha)

So the result formula should be:

P(W,alpha|y,X) is proportional to P(y|X,W,alpha) \* P(W,alpha)

I believe here that the X cannot be dismissed, because the meaning of likelihood describes the plausibility of a model parameter value, given specific observed data.

And also, The problem from last week, I insist on the opinion that OLS should be used as varification in our model. the result of performing Bayesian Linear Regression is a distribution of possible model parameters based on the data and the prior. This allows us to quantify our uncertainty about the model: if we have fewer data points, the posterior distribution will be more spread out.

So point is that as the amount of data points increases, the likelihood overcomes the prior, and in the case of infinite data, the outputs for the parameters converge to the values obtained from OLS.