**Priors:**

* Why have we been using a prior of 0 for all of our parameters thus far?
  + How do we determine what a good prior is?
* Can we just “follow our nose” when it comes to modeling this as a binary classification problem?
  + The Uncertainty classification problem suggests we can just apply softmax to the output of the network