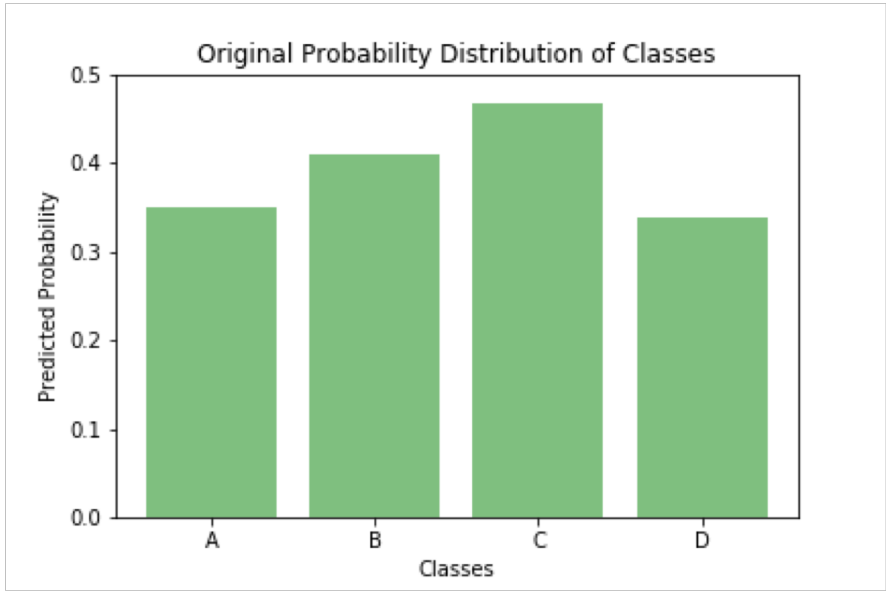
Sampling with Diversity

Uses “temperature” to alter the probability distribution of the potential tokens that we randomly sample from to allow for more (or less) stochasticity in model predictions.

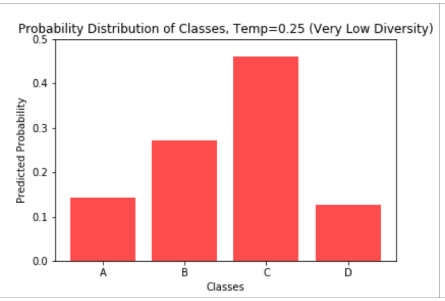
$$temperature \begin{cases} > 1, more\ diversity \\ = 1, no\ diversity \\ < 1, less\ diversity \end{cases}$$

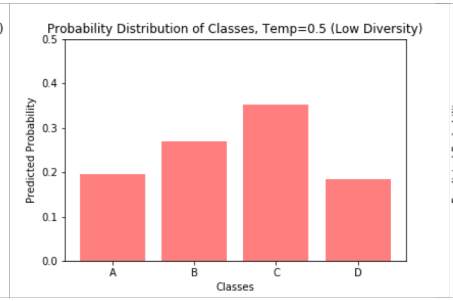


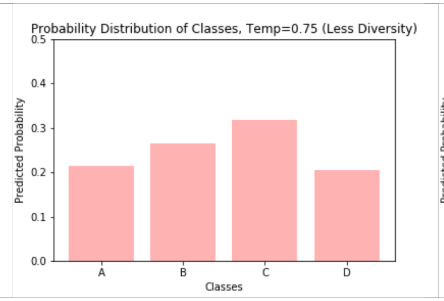
The array of predicted probabilities for each class then becomes:

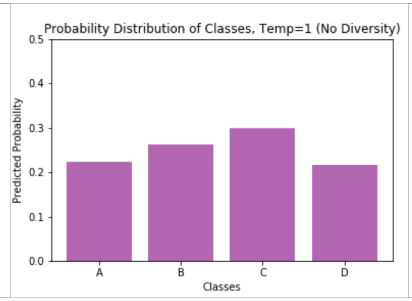
$$p_i = \frac{\exp(\frac{\widehat{probas}}{temperature})}{\sum \exp(\frac{\widehat{probas}}{temperature})}$$

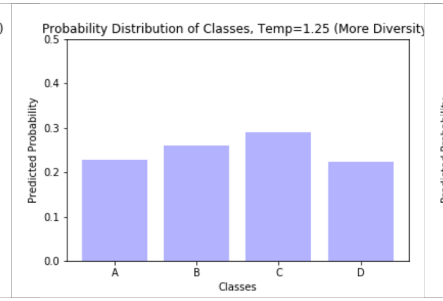
Temperature

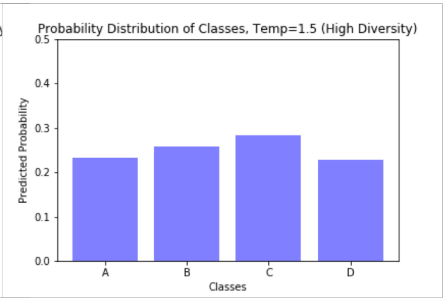












= 0.25 **= 0.5** **= 0.75** **= 1** **= 1.25** **= 1.5**

The most probable becomes even more probable

Distribution Unchanged

The least probable becomes more probable