

Question 2

(a) Given the probability that the question j is correctly answered by student i is:

$$p_{ij} = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

The log-likelihood for all students is derived as follows:

$$\begin{aligned} \log p(\mathbf{C}|\boldsymbol{\theta}, \boldsymbol{\beta}) &= \sum_{i,j} (c_{ij} \log p_{ij} + (1 - c_{ij}) \log(1 - p_{ij})) \\ &= \sum_{i=1}^n \sum_{j=1}^m \left(c_{ij} \log \left(\frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \right) + (1 - c_{ij}) \log \left(1 - \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \right) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m (c_{ij}(\theta_i - \beta_j) - \log(1 + \exp(\theta_i - \beta_j))), \end{aligned}$$

where c_{ij} is the binary response of student i to question j .

The log-likelihood with respect to θ_i is:

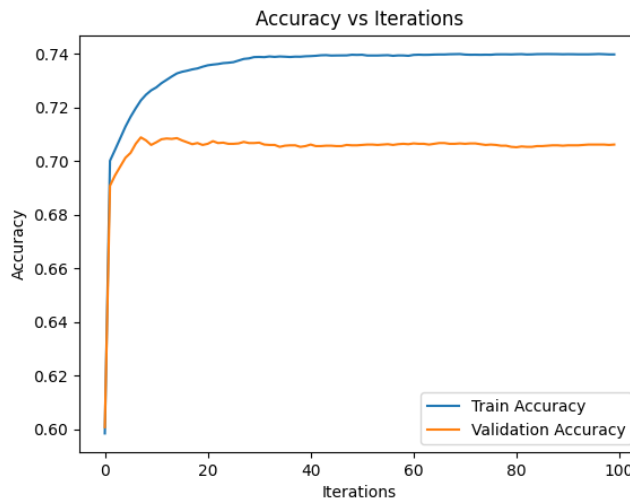
$$\begin{aligned} \frac{\partial \log p(\mathbf{C}|\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \theta_i} &= \sum_{j=1}^m \left(c_{ij} - \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \right) \\ &= \sum_{j=1}^m (c_{ij} - p_{ij}). \end{aligned}$$

The log-likelihood with respect to β_j is:

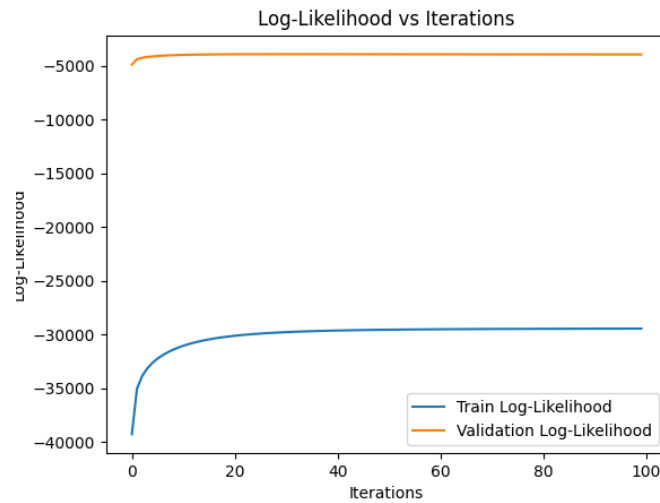
$$\begin{aligned} \frac{\partial \log p(\mathbf{C}|\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \beta_j} &= \sum_{i=1}^n \left(c_{ij} - \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \right) \\ &= \sum_{i=1}^n (c_{ij} - p_{ij}). \end{aligned}$$

(b) The hyperparameters I selected are: learning rate = 0.01 and iterations = 100.

The training and validation accuracies vs iterations are in the graph below:



The log-likelihoods vs iterations are in the graph below:

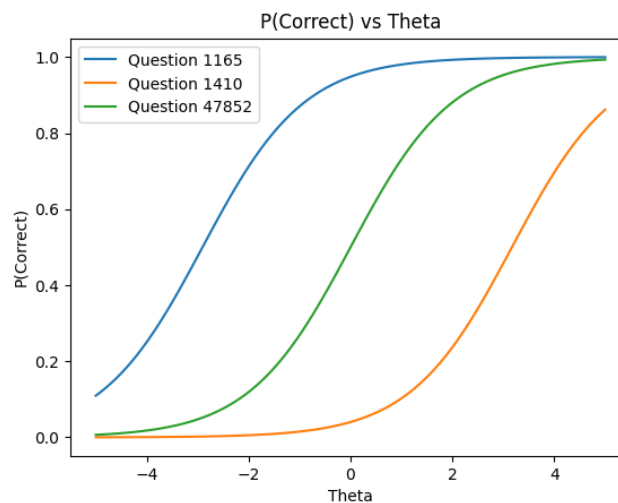


(c) The Final Validation Accuracy: 0.7063223257126728

The Final Test Accuracy: 0.707310189105278

(d) I select the lowest difficulty question j_1 (Question 1165), the highest difficulty question j_2 (Question 47852) and the average difficulty question j_3 (Question 1410).

The probability of the correct response is in the graph below:



(e) The shape of the curves are like the sigmoid function as expected.

Fix a question j . As θ_i increases, the probability of the correct response p_{ij} increases. This means if a student has a higher ability, the probability of the correct response increases.

Fix a student i . As β_j increases, the probability of the correct response p_{ij} decreases. This means if a question has a higher difficulty, the probability of the correct response decreases.