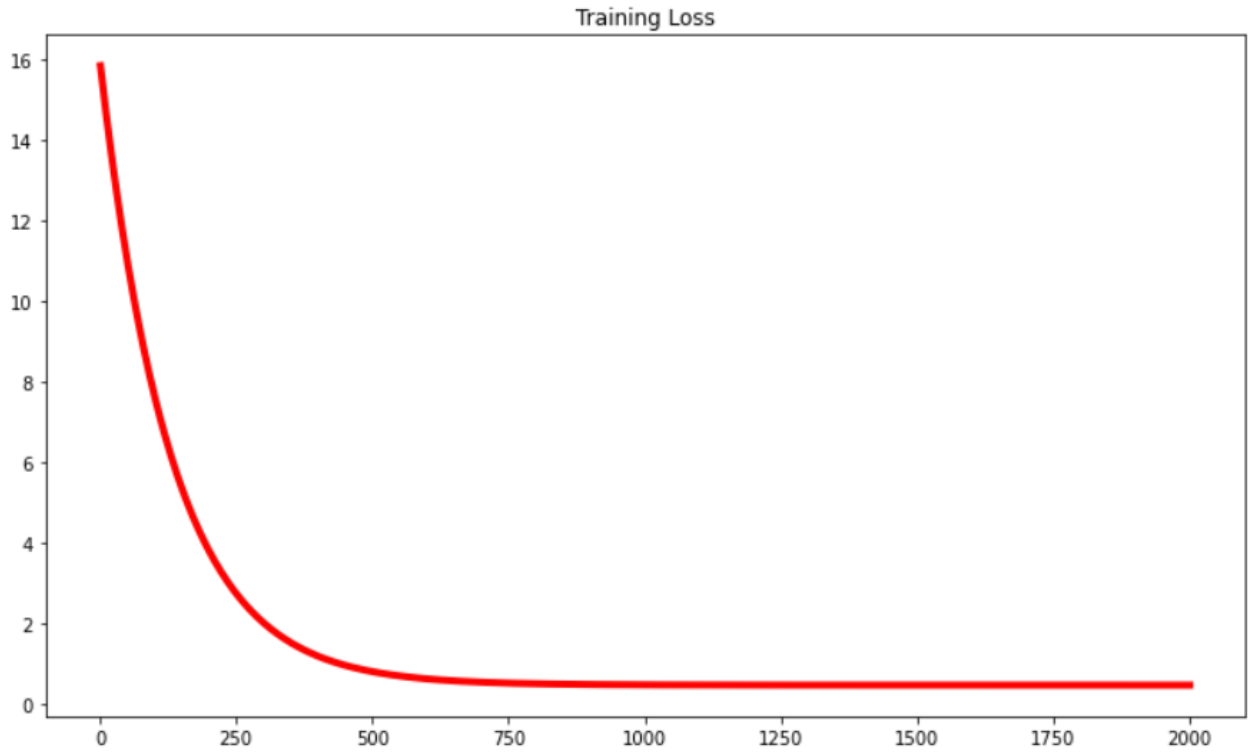
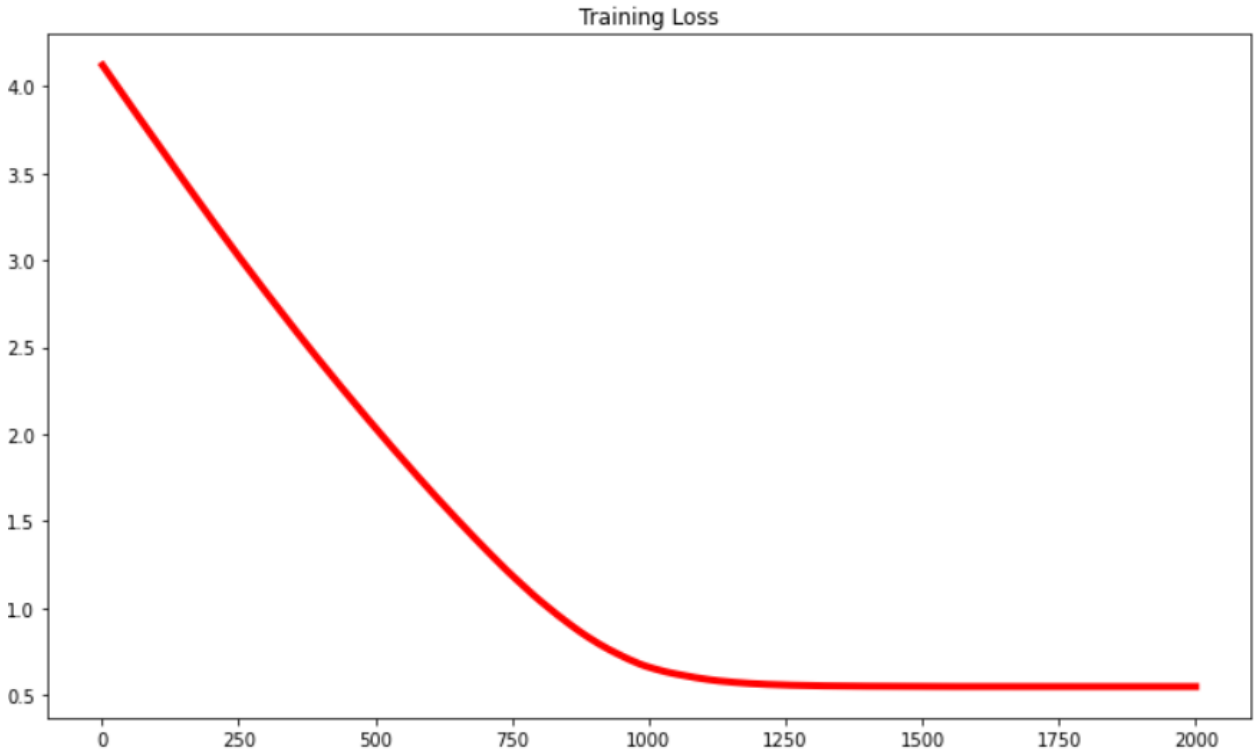
Part. 1

2. Learning curve

Objective Function MSE



Objective Function MAE



3. (mean square error, mean absolute error)

Objective Function MSE: (0.4933055707242924, 0.5625219139338563)

Objective Function MAE: (0.4943236609451898, 0.5640064991580637)

4. (weights, intercepts)

Objective Function MSE: (0.4527677032784589, -0.04147920939334472)

Objective Function MAE: (0.4338474404673587, -0.03764348286956803)

5. gradient descent 每更新一次 weights, intercepts 就要跑完所有的訓練資料；mini-batch gradient descent 把所有訓練資料分成很多份，每次更新 weights, intercepts 就只要跑完其中一份；stochastic gradient descent 跟 mini-batch gradient descent 一樣是分小份的，但是每次的樣本是隨機抽取

Part. 2

1.

① $\frac{1}{5} \times \frac{3}{10} + \frac{2}{5} \times \frac{2}{4} + \frac{2}{5} \times \frac{4}{20} = \frac{3}{50} + \frac{10}{50} + \frac{4}{50} = \frac{17}{50}$ #

② $\dfrac{\frac{2}{5} \times \frac{2}{4}}{\frac{1}{5} \times \frac{3}{10} + \frac{2}{5} \times \frac{2}{4} + \frac{2}{5} \times \frac{12}{20}} = \dfrac{\frac{10}{50}}{\frac{25}{50}} = \frac{10}{25} = \frac{2}{5}$ #

2.

$$\begin{aligned} \mathrm{Var}[f] &= E[(f(x) - E[f(x)])^2] \\ &= E[f(x)^2 - 2f(x)E[f(x)] + E[f(x)]^2] \\ &= E[f(x)^2] - 2E[f(x)]E[f(x)] + E[f(x)]^2 \\ &= E[f(x)^2] - E[f(x)]^2 \end{aligned}$$

3.

$$\begin{aligned} E_y[E_x[x \mid y]] &= E_y\left[\sum_x x \cdot P(X = x \mid Y)\right] \\ &= \sum_y \left[\sum_x x \cdot P(X = x \mid Y = y)\right] P(Y = y) \\ &= \sum_y \sum_x x \cdot P(X = x \mid Y = y) \\ &= \sum_x x \sum_y P(X = x \mid Y = y) \\ &= \sum_x x \, P(X = x) \\ &= E(X) \end{aligned}$$