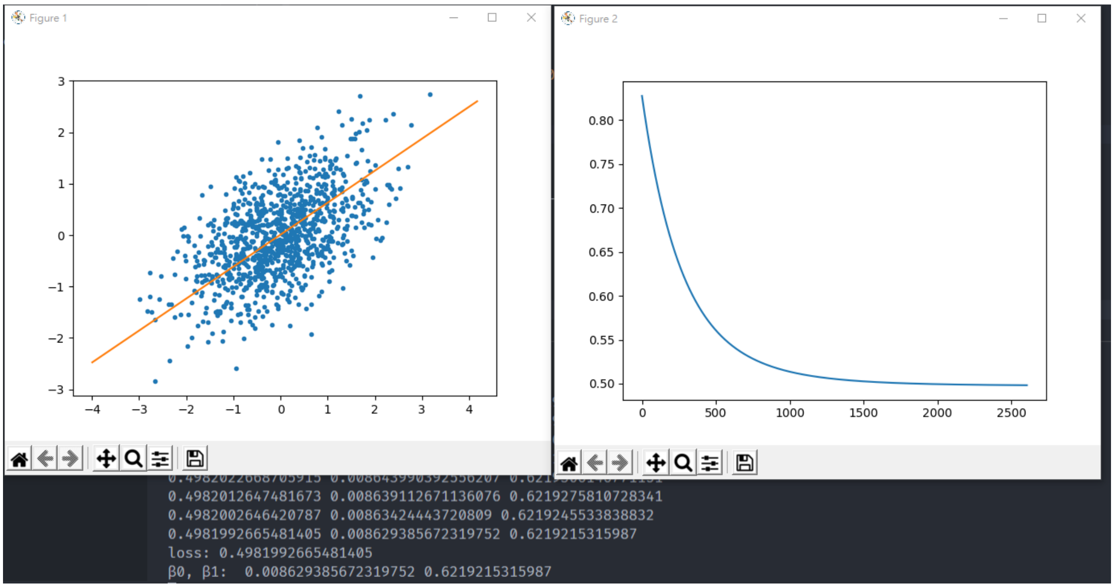
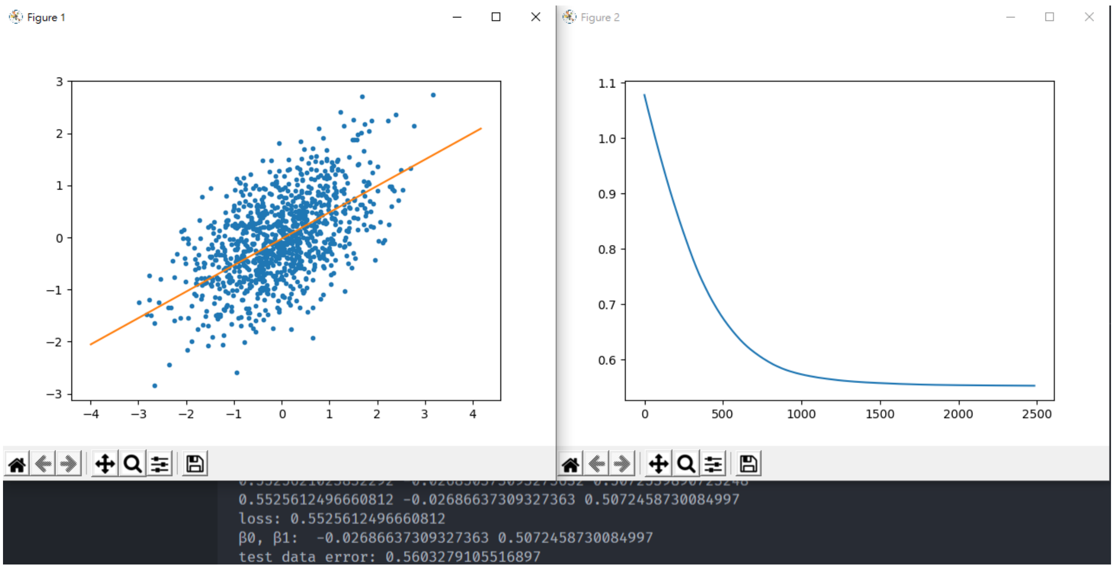
Pattern Recognition, Homework 1

1. Part 1
   (1) Implement the linear regression model
   (2) Plot the learning curve of the training with both losses

MSE



MAE



(3) mean square error and mean absolute error between your predictions and the ground truths on the testing data

MSE	MAE
0.5098033892530457	0.5603279105516897

(4) the weights (β1) and intercepts (β0) of your linear model trained from both losses?

	MSE	MAE
β0	0.008629385672319752	- 0.02686637309327363
β1	0.6219215315987	0.5072458730084997

(5) Please explain the difference between gradient descent, mini-batch gradient descent, and stochastic gradient descent?

- gradient descent:
  一次疊代(更新 weight) 就需要跑完全部的訓練資料。
  耗時較久，不過也較保險，可以避免只取到較邊緣的訓練資料導致訓練方向錯誤。
- mini-batch gradient descent:
  將資料集分成幾堆，一次疊代只需要跑完一堆就對 weight 進行更新。
  上下兩者的平衡。
- stochastic gradient descent:
  將資料分堆到極致，一次疊代只需要跑完一**筆**資料就對 weight 進行更新。
  收斂速度快，不過 weight 更新方向卻容易跳動(因為訓練資料大部分都帶有誤差，只使用一筆資料訓練會將誤差對訓練方向的影響加到最大)。

2. Part 2

(1)

① $$P(\text{Guava}) = P(\text{Guava} \mid box=R) \times P(box=R) + P(\text{Guava} \mid box=B) \times P(box=B) + P(\text{Guava} \mid box=G) \times P(box=G)$$

$$= (3/10)\times 0.2 + (2/4)\times 0.4 + (4/20)\times 0.4$$

$$= 0.06 + 0.2 + 0.08$$

$$= 0.34 \ \#$$

② $$P(box=B \mid \text{Apple}) = P(box=B \,\&\, \text{Apple}) / P(\text{Apple})$$

$$= 0.4\times\frac{2}{4} \Big/ \left(0.2\times\frac{3}{10} + 0.4\times\frac{2}{4} + 0.4\times\frac{12}{20}\right)$$

$$= 0.2 / (0.06 + 0.2 + 0.24)$$

$$= 0.2 / 0.5$$

$$= 0.4 \ \#$$

(2)

$$var(f) = E\left[\left[f(x)-E(f(x))\right]^2\right] = E\left[f(x)^2 - 2f(x)\cdot E[f(x)]\cdot E[f(x)]^2\right]$$

$$= E\left[f(x)^2\right] \quad -2E\left[f(x)E[f(x)]\right] \quad +E\left[E[f(x)]^2\right]$$

$$= E[f(x)^2] \quad -2E[f(x)]E[f(x)] \quad +E[f(x)]^2$$

$$= E[f(x)^2] - E[f(x)]^2 \ \#$$

(3)

$$E_y\left[E_x[x|y]\right]$$

$$= E_y\left[\sum_{x \in X} P(X=x|Y=y)\cdot x\right]$$

$$= \sum_{y \in Y} P(Y=y)\left[\sum_{x \in X} P(X=x|Y=y)\cdot x\right]$$

$$= \sum_{y \in Y}\sum_{x \in X} P(Y=y)P(X=x|Y=y)\cdot x$$

$$= \sum_{y \in Y}\sum_{x \in X} P(X=x \ \& \ Y=y)\cdot x$$

$$= \sum_{x \in X} P(X=x)\cdot x$$

$$= E[x]$$