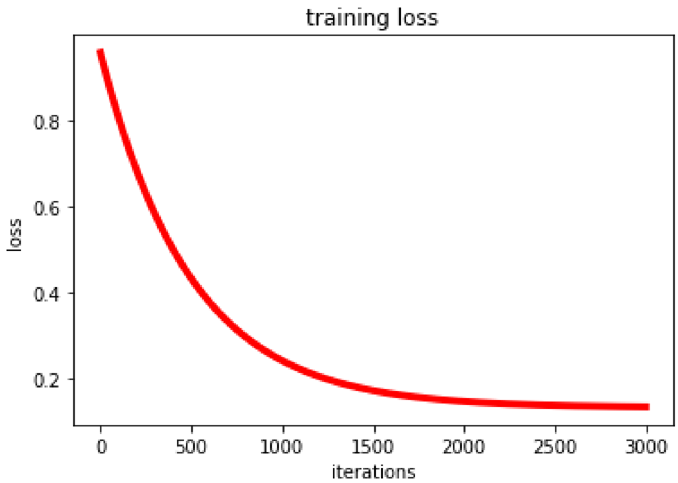
NYCU Pattern Recognition, HW1

310551079 梁友誠

Part. 1, Coding:

1. The learning curve of the training.



2. Mean Square Error of prediction and ground truth is:

$$\boldsymbol{Mean((y\_test - prediction)^2) = 0.07662689789 62181}$$

```
# Q2. Mean Square Error of your prediction and ground truth
mse = np.mean((y_pred-y_test)**2)
print(f'MSE: {mse}')
```

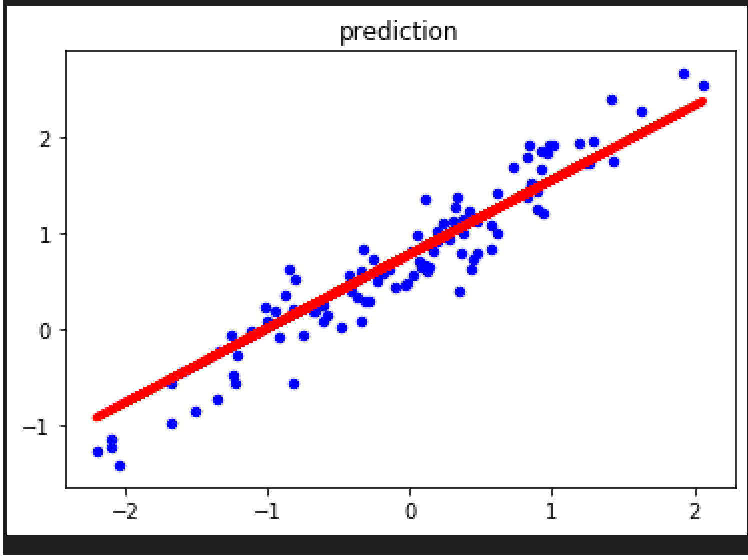
```
MSE: 0.0766268978962181
```

3. The weights and intercepts of my linear model:

```
print(f'intercepts(w0): {w0}')
print(f'weights(w1): {w1}')

# Visualize prediction with y_test
pred_fig(x_test, y_test, y_pred, "prediction")
```

```
intercepts(w0): 0.7413464799473299
weights(w1): 0.7741365407248483
```



4. The main difference between Gradient Descent, Mini-Batch Gradient Descent, and Stochastic Gradient Descent is the **frequency of updating parameters**.
   In Gradient Descent, **all the training data is taken into consideration to take a single step**, so that’s just one step of gradient descent in one epoch.
   In Stochastic Gradient Descent, we **consider just one training data at a time to update the parameters**. So in one iteration, it will update the weights by *n* times (*n* refers to the number of training data).
   In Mini-Batch Gradient Descent, we use a batch of a fixed number of training data which is less than the actual dataset and call it a mini-batch, and then use mini-batch to update weights. So in one iteration, the model will calculate the mean gradient of the mini-batch to update the weights. For example, if we have 100 training data and divide it into 10 batches (10 data for each batch), the model will use these 10 batches to update weights by 10 times in one iteration.

Part. 2, Questions:

Question 1

$$R \begin{cases} 3 \text{ apples} \\ 4 \text{ oranges} \\ 3 \text{ guavas} \end{cases} \quad 0.2$$

$$B \begin{cases} 2 \text{ apples} \\ 0 \text{ orange} \\ 2 \text{ guavas} \end{cases} \quad 0.4$$

$$G \begin{cases} 12 \text{ apples} \\ 4 \text{ oranges} \\ 4 \text{ guavas} \end{cases} \quad 0.4$$

(1) The probability of selecting guava:

$$\frac{1}{5} \times \frac{3}{10} + \frac{2}{5} \times \frac{2}{4} + \frac{2}{5} \times \frac{4}{20}$$

$$= \frac{17}{50} \quad \#$$

(2) The probability of selecting apple that came from blue box:

$$\frac{\frac{2}{5} \times \frac{2}{4}}{\frac{1}{5} \times \frac{3}{10} + \frac{2}{5} \times \frac{2}{4} + \frac{2}{5} \times \frac{12}{20}}$$

$$= \frac{\frac{10}{50}}{\frac{25}{50}}$$

$$= \frac{2}{5} \quad \#$$

Question 2 :

Let R_1 be the distribution are of class C_1

.. R_2 ' - - . - . - ' ' . . - . . C_2

$$\Rightarrow p(\text{mistake}) = \int_{R_1} p(x, C_2)\,dx + \int_{R_2} p(x, C_1)\,dx$$

In error made in R_1 always have $p(C_1|x) \geq p(C_2|x)$

and we know if $a \leq b$, $a \leq (ab)^{\frac{1}{2}}$

we can get $p(C_2|x) \leq \{p(C_1|x)\,p(C_2|x)\}^{\frac{1}{2}}$

$$\Rightarrow \int_{R_1} p(x, w_2) = \int_{R_1} p(C_2|x)\,p(x)\,dx \quad \text{... product rule}$$
$$\leq \int_{R_1} \{p(C_1|x)\,p(C_2|x)\}^{\frac{1}{2}} p(x)\,dx$$
$$= \int_{R_1} \{p(x, C_1)\,p(x, C_2)\}^{\frac{1}{2}}\,dx$$

similar situations apply R_2, we get $p(C_2|x) \geq p(C_1|x)$

$$\Rightarrow \int_{R_2} p(x, w_1) = \int_{R_2} p(C_1|x)\,p(x)\,dx$$
$$= \int_{R_2} \{p(x, C_1)\,p(x, C_2)\}^{\frac{1}{2}}\,dx$$

So, $p(\text{mistake}) = \int_{R_1} p(x, C_2)\,dx + \int_{R_2} p(x, C_1)\,dx$

$$= \int_{R_1} \{p(x, C_1)\,p(x, C_2)\}^{\frac{1}{2}}\,dx + \int_{R_2} \{p(x, C_1)\,p(x, C_2)\}^{\frac{1}{2}}\,dx$$
$$= \int \{p(x, C_1)\,p(x, C_2)\}^{\frac{1}{2}}\,dx$$

#

Question 3.

(1) we know $E_x[x|y] = \int p(x|y) f(x) dx$

and $E_y(x) = \int p(y) f(y) dy$

so we get $E_y[E_x(x|y)]$

$= E_y[\int p(x|y) f(x) dx]$

$= \iint p(y)\, p(x|y)\, f(x)\, dx\, dy$

$= \iint \underbrace{p(x|y)\, p(y)}_{\text{product rule}} f(x)\, dx\, dy$

$= \iint \underline{p(x,y)} f(x)\, dx\, \underbrace{dy}_{\text{sum rule}}$

$= \int p(x) f(x)\, dx$

$= E[x]$ #

(2) By $\mathrm{var}(x) = E[(x - E(x))^2] = E[x^2] - (E(x))^2$

we can derive

$E_y[\mathrm{var}_x[x|y]] = E_y[E_x[(x - E_x[x|y])^2 | y]]$

$= \underline{E_y[(E_x(x|y))^2]} - (E_y[E_x[x|y]])^2$

$\mathrm{var}_y[E_x(x|y)] = E_y[(E_x(x|y) - E_y[E_x(x|y)])^2]$

$= E_y[E_x[x^2|y]] \underline{- E_y[(E_x(x|y))^2]}$

we also know that $E_y[E_x(x|y)] = E(x)$ from part (1)

so we can get $E_y[\mathrm{var}_x(x|y)] + \mathrm{var}_y[E_x(x|y)]$

$= E_y[E_x(x^2|y)] - (E_y[E_x(x|y)])^2$

$= E[x^2] - E[x]$

$= \mathrm{var}(x)$ #