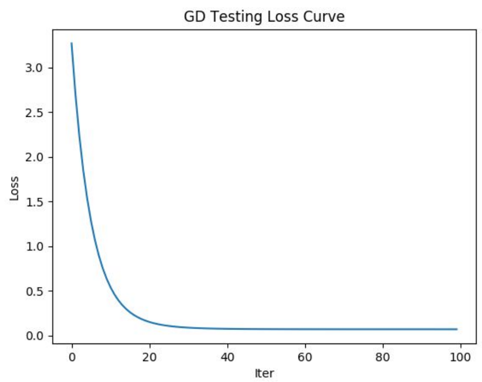
Pattern Recognition, Homework 1

0516003 李智嘉

Part. 1, Coding:

1. Learning Curve
    - Training loss curve

    - Testing loss curve

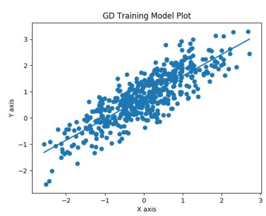


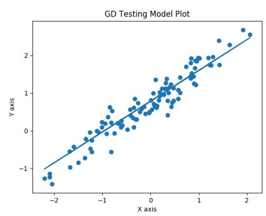
2. Mean Square Error
    - Training MSE: 0.5347
    - Testing MSE: 0.0687

```
Training loss 0.5347280122225704, Testing loss 0.06869637127070814
```

3. Weights and Intercepts
    - Weights: 0.8180
    - Intercepts: 0.7845

```
Model Weight and Intercepts: [[0.8179535 0.7845594]]
```





4. Difference between GD, Mini-Batch GD, SGD
    - Gradient Descent

- feed all the training data into model at one time, and compute the gradient, then doing backpropagation.

- Mini-Batch Gradient Descent
  - divide the training data into some batches, and feed one batch at one time, then compute gradient and doing backpropagation.
- Stochastic Gradient Descent
  - feed one data into the model at one time, just like mini-batch gradient descent with batch size 1.

Part. 2, Questions:

1.

$$
\begin{aligned}
p(\text{guava}) &= p(R)\,p(\text{guava}|R) + p(G)\,p(\text{guava}|G) \\
&\quad + p(B)\,p(\text{guava}|B) \\
&= 0.2\cdot\frac{3}{10} + 0.4\cdot\frac{4}{20} + 0.4\cdot\frac{2}{4} \\
&= 0.06 + 0.08 + 0.2 \\
&= \underline{\underline{0.34}}\ \#
\end{aligned}
$$

$$
\begin{aligned}
p(B|\text{apple}) &= \frac{p(\text{blue}, \text{apple})}{p(\text{apple})} = \frac{p(\text{apple}|B)\,p(B)}{p(\text{apple})} \\
&= \frac{0.4\cdot\frac{2}{4}}{p(R)p(\text{apple}|R) + p(G)p(\text{apple}|G) + p(B)p(\text{apple}|B)} \\
&= \frac{0.2}{0.2\cdot\frac{3}{10} + 0.4\cdot\frac{12}{20} + 0.4\cdot\frac{2}{4}} = \frac{0.2}{0.06+0.24+0.2} \\
&= \frac{0.2}{0.5} = \underline{\underline{0.4}}\ \#
\end{aligned}
$$

2.

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