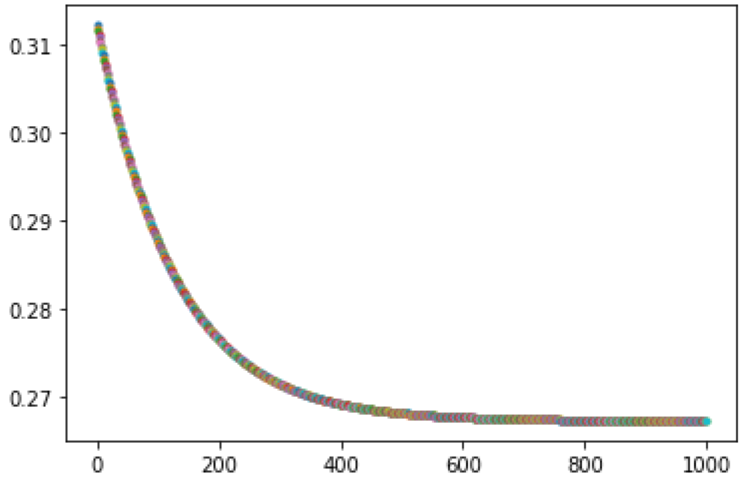
Part I

1.



2.

MSE: 0.069171

3.

weight: 0.816531, intecept: 0.788410

4. Gradient Descent uses all of the data to do one iteration of updating weights;

Mini-batch Gradient Descent uses a subset of data to update; while Stochastic Gradient Descent uses only one data to update.

5.

```
def model(x_test):
    var = np.var(x_train)
    # initialize the starting value with normal r.v
    m, c = np.random.normal(0, var), np.random.normal(0, var)
    learning_rate = 2e-3
    iteration = 1000
    for i in range(iteration):
        n = x_train.size

        # prediction list of y
        y = (m*x_train + c)

        # loss function determined by MSE
        loss = np.sum((y-y_train)**2) / n

        # plot iteration-loss graph
        plt.plot(i, loss, '.')

        # gradient of m after partial derivative
        gradient_m = (2/n) * np.sum(x_train*(y-y_train))

        # gradient of c after partial derivative
        gradient_c = (2/n) * np.sum(y-y_train)

        # renew m
        m -= learning_rate * gradient_m

        # renew c
        c -= learning_rate * gradient_c
    print("weight: %f, intecept: %f" %(m, c))
    return m*x_test + c
```

```
# MSE of prediction and ground truth
mse = np.sum((y_pred-y_test)**2) / x_test.size
print("MSE: %f" %(mse))
```

Part II

1. (a) $0.2 \times \frac{3}{10} + 0.4 \times \frac{2}{4} + 0.4 \times \frac{4}{20} = 0.34$ #

(b) $$\frac{0.4 \times \frac{2}{4}}{0.2 \times \frac{3}{10} + 0.4 \times \frac{2}{4} + 0.4 \times \frac{12}{20}} = \frac{2}{5} = 0.4$$ #

2. $a = a^{\frac{1}{2}} \cdot a^{\frac{1}{2}} \leq a^{\frac{1}{2}} \cdot b^{\frac{1}{2}}$ ($\because a < b$)

$= (ab)^{\frac{1}{2}}$ #

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

To minimize the misclassification rate, $p(x, C_1) \geq p(x, C_2)$ when $x \in R_1$

$p(x, C_2) \geq p(x, C_1)$ when $x \in R_2$

$$\Rightarrow p(\text{mistake}) \leq \int_{R_1} \{p(x, C_1)p(x, C_2)\}^{\frac{1}{2}} dx + \int_{R_2} \{p(x, C_1)p(x, C_2)\}^{\frac{1}{2}} dx$$

$$= \int \{p(x, C_1)p(x, C_2)\}^{\frac{1}{2}} dx$$ #

3. $E[X] = \int_{x\in X} x \cdot p_X(x)\,dx = \int_{x\in X} x \cdot \int_{y\in Y} p_{XY}(x,y)\,dy\,dx$

$= \int_{y\in Y}\int_{x\in X} x \cdot [p_{X|Y}(x|y) \cdot p_Y(y)]\,dx\,dy$

$= \int_{y\in Y}\left(\int_{x\in X} x \cdot p_{X|Y}(x|y)\,dx\right) p_Y(y)\,dy$

$= \int_{y\in Y} E_{X|Y}[x|y] \cdot p_Y(y)\,dy$

$= E_Y[E_{X|Y}[x|y]] \Rightarrow E_y[E_x[x|y]]$ #

$Var[x] = E[x^2] - E^2[x] = E_Y[E_x[x^2|Y]] - E_Y^2[E_x[x|Y]]$

$= E_Y[E_x[x^2|Y]] + (Var_Y[E_x[X|Y]] - E_Y[E_x^2[x|Y]])$

$= E_Y[E_x[x^2|Y] - E_x^2[x|Y]] + Var_Y[E_x[X|Y]]$

$= E_Y[Var_x[X|Y]] + Var_Y[E_x[X|Y]]$

$\Rightarrow E_y[Var_x[x|y]] + Var_Y[E_x[x|y]]$ #