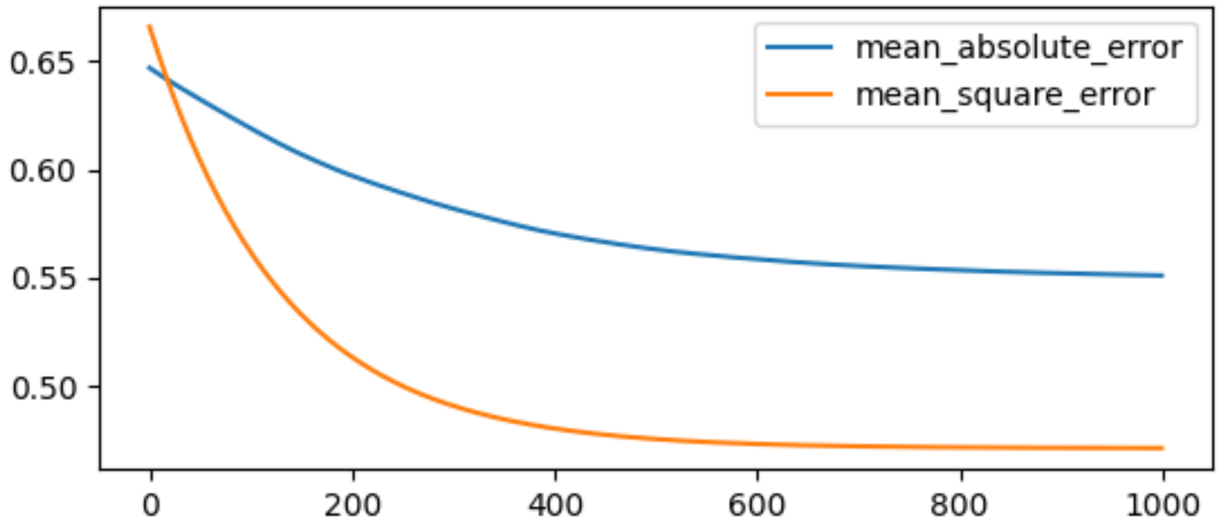
資工4B 游騰德 0616026

Pattern Recognition

HW1: Regression

Part 1, Coding

Learning curve of the training with both losses



Error between predictions and the ground truths on the testing data

---

Mean square error

0.4917001061329206

Mean absolute error

0.56824647542189

Weights (β1) and intercepts (β0)

Mean square error

$\beta 1 = 0.4434104122525014$
$\beta 0 = -0.0016105435982399672$

Mean absolute error

β1 = 0.3874517678316522
β0 = −0.026438438438438714

Gradient descent, mini-batch gradient descent, and stochastic gradient descent

---

Gradient descent

Calculate the gradients on each observation one by one

Mini-batch gradient descent

Calculate the gradients for a group of observations rather than for each observation which results in a faster optimization

---

Stochastic gradient descent

Chose the random observations randomly and calculate the gradients

Part 2, Questions

Question 1

P(selecting a guava)

$= 0.2 \times 0.3 + 0.4 \times 0.5 + 0.4 \times 0.2$

$= 0.34$ ※

P(from Box B | selected apple)

$$= \frac{0.4 \times 0.5}{0.2 \times 0.3 + 0.4 \times 0.5 + 0.4 \times 0.6}$$

$= 0.4$ ※

Question 2

$$
\begin{aligned}
\mathrm{var}[f] &= E[(f(x) - E[f(x)])^2] \\
&= E[f(x)^2 - 2f(x)E[f(x)] + E[f(x)]^2] \\
&= \int [f(x)^2 - 2f(x)E[f(x)] + E[f(x)]^2] \, p(x)dx \\
&= \int f(x)^2 p(x)dx - \int 2f(x)E[f(x)] \, p(x)dx + \\
&\quad \int E[f(x)]^2 p(x)d(x) \\
&= E[f(x)^2] - 2E[f(x)]E[f(x)] + E[f(x)]^2 \\
&= E[f(x)^2] - E[f(x)]^2
\end{aligned}
$$

Question 3

$$E_y[E_x[x|y]]$$

$$= \int E_x[x|y] f_y(y)\,dy$$

$$= \int \left(\int x f_{x|y}(x|y)\,dx \right) f_y(y)\,dy$$

$$= \iint x f_{x|y}(x|y) f_y(y)\,dx\,dy$$

$$= \iint x f_{x,y}(x,y)\,dx\,dy$$

$$= \iint x f_{x,y}(x,y)\,dy\,dx$$

$$= \int x \left(\int f_{x,y}(x,y)\,dy \right) dx$$

$$= \int x f_x(x)\,dx$$

$$= E[x]$$