CSE512 Fall 2018 - Machine Learning - Homework 3

Name: Manideep Attanti
Solar Id: 112028167
Netid email: manideep.attanti@stonybrook.edu

Q1 1.1

(1.1) $X = (x_1, x_2)$

$x_1 \rightarrow$ boolean (bernouli)

$x_2 \rightarrow$ continuous (assuming gaussian)

we have to calculate 7 parameters to calculate $P(Y|X)$

for $x_1 \rightarrow$ 2 params

$P(x_1 = 0 | Y = 0)$ and

$P(x_1 = 0 | Y = 1)$

for $x_2 \rightarrow$ 4 params

μ_0, σ_0 which are mean and standard deviation when $Y = 0$ &

μ_1, σ_1 when $Y = 1$

$P(Y=0) \rightarrow$ 1 param which is the prior.

$\Rightarrow$ Calculating $P(Y=1|X)$ will suffice

$\therefore P(Y=0|X) = 1 - P(Y=1|X)$

$$P(Y=1|X) = \frac{P(X|Y=1) \cdot P(Y=1)}{P(X)}$$

$$= \frac{P(x_1|Y=1) \cdot P(x_2|Y=1) \cdot P(Y=1)}{P(x_1|Y=1) \cdot P(x_2|Y=1) \cdot P(Y=1) + P(x_1|Y=0) \cdot P(x_2|Y=0) \cdot P(Y=0)}$$

Dividing by numerator

$$P(Y=1|X) = \frac{1}{1 + \frac{P(Y=0)}{P(Y=1)} \cdot \frac{P(x_1|Y=0) \cdot P(x_2|Y=0)}{P(x_1|Y=1) \cdot P(x_2|Y=1)}}$$

$$P(x_1|Y=0) = P(x_1=1|Y=0)^{x_1} \cdot P(x_1=0|Y=0)^{1-x_1}$$

$$P(x_1|Y=1) = P(x_1=1|Y=1)^{x_1} \cdot P(x_1=0|Y=1)^{1-x_1}$$

$$P(x_2|Y=0) = \frac{1}{\sqrt{2\pi}\,\sigma_0^2} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}}$$

$$P(x_2|Y=1) = \frac{1}{\sqrt{2\pi}\,\sigma_1^2} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

Let, $P(Y=0) = k \Rightarrow P(Y=1) = 1-k$

$P(x_1=1|Y=0) = a \Rightarrow P(x_1=0|Y=0) = 1-a$

$P(x_1=1|Y=1) = b \Rightarrow P(x_1=0|Y=1) = 1-b$

Substituting in above eq,n. we get

$$P(Y=1|x) = \frac{1}{1 + \frac{k}{1-k} \cdot \frac{a^{x_1} \cdot (1-a)^{1-x_1}}{b^{x_1} \cdot (1-b)^{1-x_1}} \cdot \frac{\sigma_1^2}{\sigma_0^2} e^{-\left[\frac{(x-\mu_0)^2}{2\sigma_0^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right]}}$$

The params are $k = P(Y=0)$

$a = P(x_1=1|Y=0)$

$b = P(x_1=1|Y=1)$

$\mu_0, \mu_1, \sigma_0, \sigma_1$

Q1. 1.2

1.2 $Y \rightarrow$ boolean

$X = (x_1, x_2, \ldots, x_d)$ vector of booleans

$$P(Y=1|X) = \frac{P(X|Y=1) \cdot P(Y=1)}{P(X|Y=1) \cdot P(Y=1) + P(X|Y=0) \cdot P(Y=0)}$$

dividing by numerator

$$= \frac{1}{1 + \frac{P(Y=0)}{P(Y=1)} \cdot \frac{P(X|Y=0)}{P(X|Y=1)}}$$

$$P(X|Y=0) = \prod_{i=1}^{d} P(x_i|Y=0)$$

$$P(x_i|Y=0) = P(x_i=1|Y=0)^{x_i} \cdot P(x_i=0|Y=0)^{1-x_i}$$

$\uparrow$ (Bernouli)

Let $P(Y=0) = k$

$P(Y=1) = 1-k$

$$\frac{1}{1 + \exp\left(\ln\left(\frac{P(Y=0)}{P(Y=1)} \cdot \frac{P(X|Y=0)}{P(X|Y=1)}\right)\right)}$$

$$= \frac{1}{1 + \exp\left[\ln\left(\frac{k}{1-k} \cdot \frac{\prod_{i=1}^{d} P(x_i|Y=0)}{\prod_{i=1}^{d} P(x_i|Y=1)}\right)\right]}$$

Let's evaluate the expression inside the exponential

$$\ln k - \ln(1-k) + \sum_{i=1}^{d} \log \hat{P}(x_i|Y=0) - \sum_{i=1}^{d} \log \hat{P}(x_i|Y=1)$$

$$= \ln k - \ln(1-k) + \sum_{i=1}^{d} \left[\log(P(x_i|Y=0)) - \log(P(x_i|Y=1))\right]$$

~~$= \ln k - \ln(1-k)$~~ Evaluating the expression inside summation using previous equation

$$\log\left(P(x_i=1|Y=0)^{x_i} \cdot P(x_i=0|Y=0)^{1-x_i}\right)$$
$$- \log\left(P(x_i=1|Y=1)^{x_i} \cdot P(x_i=0|Y=1)^{(1-x_i)}\right)$$

$$= x_i \log P(x_i=1|Y=0) + (1-x_i)\cdot \log P(x_i=0|Y=0)$$
$$- x_i \log P(x_i=1|Y=1) - (1-x_i) \log P(x_i=0|Y=1)$$

$$= x_i\big(\log P(x_i=1|Y=0) - \log P(x_i=0|Y=0) +$$
$$\log P(x_i=0|Y=1) - \log P(x_i=1|Y=1)\big)$$
$$+ \left(\log P(x_i=0|Y=0) - \log P(x_i=0|Y=1)\right)$$

$$= a x_i + b$$

$\Rightarrow$ expression inside exp $= \ln k - \ln(1-k) + \sum_{i=1}^{d} (a x_i + b)$

$$= \sum_{i=1}^{d} a x_i + \ln k - \ln(1-k) + \sum_{i=1}^{d} b$$

$$= -\left(\sum_{i=1}^{d} \theta_i x_i + \theta_{d+1}\right)$$

where $\theta_i = -a$

$$= -\left[\log P(x_i=1|Y=0) + \log P(x_i=0|Y=1) - \log P(x_i=0|Y=0) - \log P(x_i=1|Y=1)\right]$$

$$\theta_{d+1} = \ln(1-k) - \ln k - \sum_{i=1}^{d} b$$

$$= \ln(1-k) + \ln k - \sum_{i=1}^{d}\left(\log P(x_i=0|Y=0) - \log P(x_i=0|Y=1)\right)$$

$$\therefore P(Y=1|x) = \frac{1}{1+\exp\left(-\left(\sum_{i=1}^{d}\theta_i x_i + \theta_{d+1}\right)\right)}$$

Hence, Logistic Regression is also the discriminative counter part to a Naive Bayes generative classifier over boolean features.

Q2. 2.1

(2.1) $\log(P(y^i \mid \bar{x}^i; \theta))$

$$= y^i \cdot \log P(y^i = 1 \mid \bar{x}^i; \theta) + (1 - y^i) \cdot \log P(y^i = 0 \mid \bar{x}^i \theta)$$

$$P(y^i = 1 \mid \bar{x}^i; \theta) = \frac{e^{\theta^T \bar{x}^i}}{1 + e^{\theta^T \bar{x}_i}}$$

$$P(y^i = 0 \mid \bar{x}^i; \theta) = \frac{1}{1 + e^{\theta^T \bar{x}_i}}$$

$\therefore$ Log likelihood $= y^i \cdot \log\left(\frac{e^{\theta^T \bar{x}^i}}{1 + e^{\theta^T \bar{x}^i}}\right)$

$$+ (1 - y^i) \cdot \log\left(\frac{1}{1 + e^{\theta^T \bar{x}^i}}\right)$$

$$= y^i \cdot \theta^T \bar{x}^i - y^i \cancel{\log}(1 + e^{\theta^T \bar{x}_i})$$

$$+ y^i \log \cancel{(1+} e^{\theta^T \bar{x}^i}) - \log(1 + e^{\theta^T \bar{x}^i})$$

$$= y^i \theta^T \bar{x}^i - \log(1 + e^{\theta^T \bar{x}^i})$$

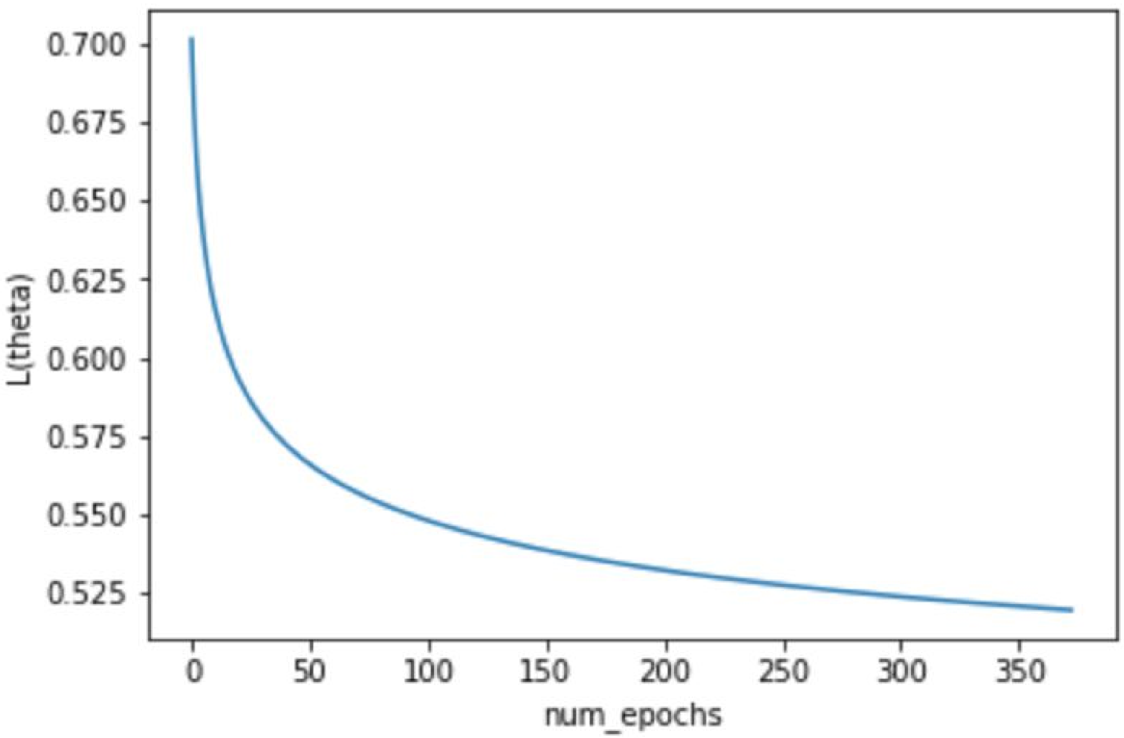
$$\frac{\partial \log(P(y^i \mid \bar{x}^i; \theta))}{\partial \theta} = y^i \bar{x}^i - \frac{1}{1 + e^{\theta^T \bar{x}^i}} \cdot e^{\theta^T \bar{x}^i} \cdot \bar{x}^i$$

$$= \left(y^i - \frac{e^{\theta^T \bar{x}^i}}{1 + e^{\theta^T \bar{x}^i}}\right) \bar{x}^i$$

$$= \left(y^i - P(y = 1 \mid \bar{x}^i; \theta)\right) \bar{x}^i$$

Q2. 2.3.1

a) Number of epochs till termination = 373
b) The graph of L(theta) and num_epochs:



c) Final value of L(θ) after optimization = 0.5196598086704581

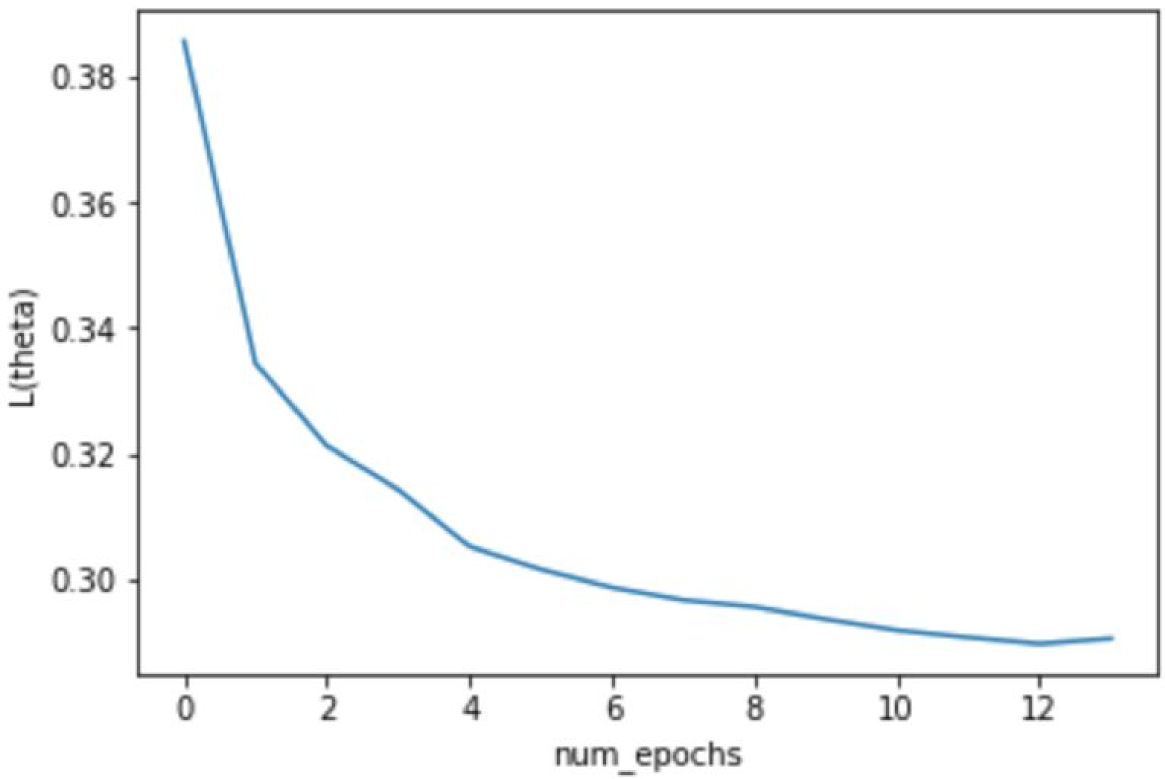
Q2 2.3.2

(a) Best value for, eta_0 = 6.4, eta_1 = 1

Number of epochs for training = 14

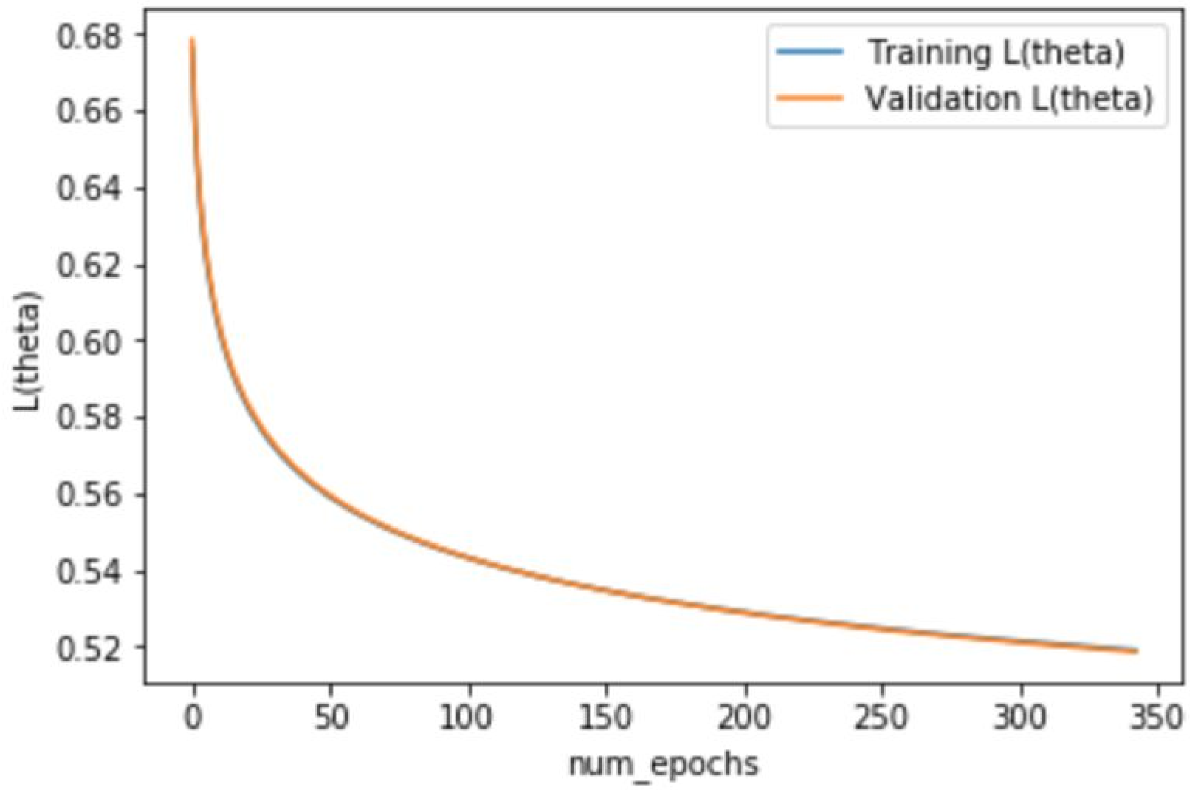
Final value of L(theta) = 0.28773104119135495

b) The graph of L(theta) and num_epochs:

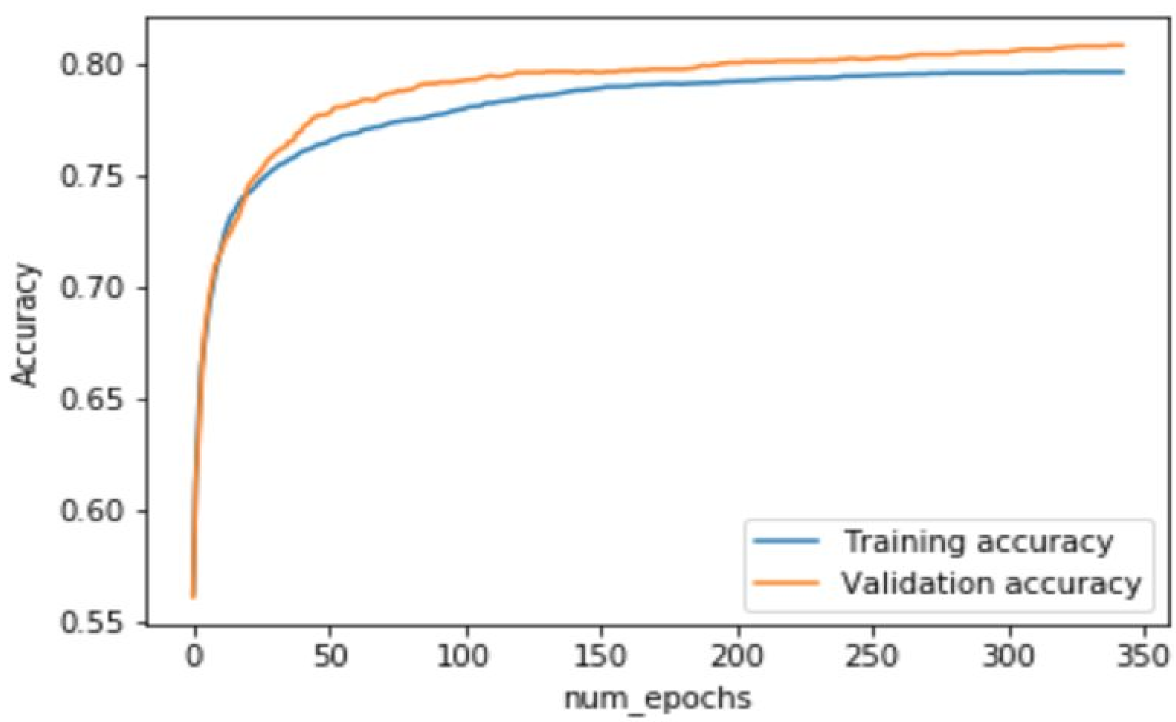

L(theta)
0.38
0.36
0.34
0.32
0.30
0
2
4
6
8
10
12
num_epochs

Q2. 2.3.3

a) The graph of L(theta) and num_epochs:

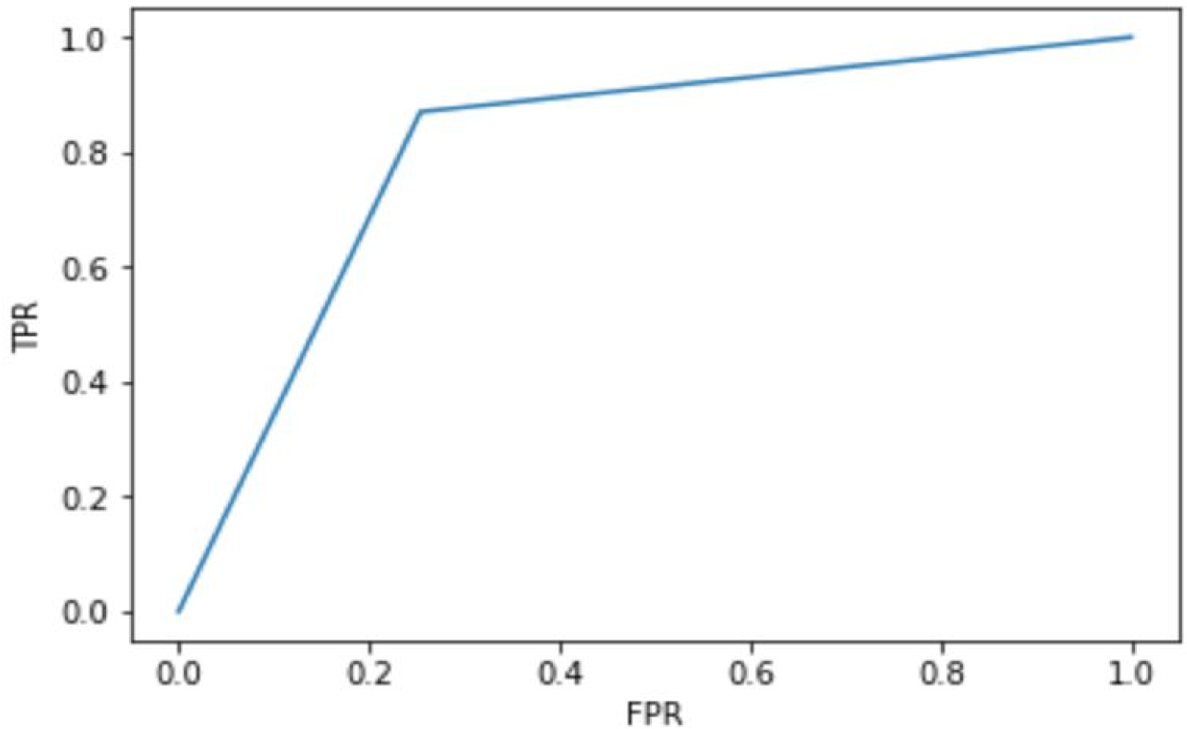


b) The graph of accuracy and num_epochs:



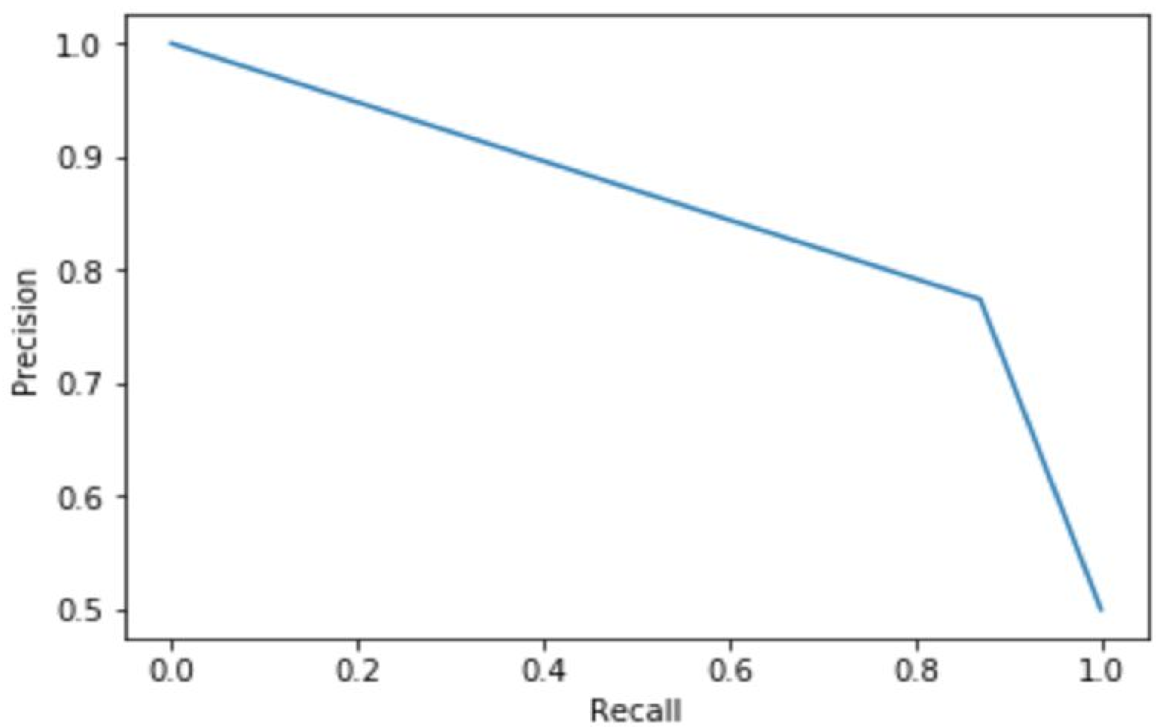
Q2.3.4

a) ROC curve on validation data.



Area under curve = 0.8080029368575624

b) Precision-Recall curve on validation data



Average Precision = 0.7383954801934758

Q2.4 Kaggle submission accuracy: 88.704%