# Cover page for answers.pdf
# CSE512 Fall 2018 - Machine Learning - Homework 3

Your Name: Sriram Reddy Kalluri

Solar ID: 111788857

NetID email address: skalluri@cs.stonybrook.edu and sriram.kalluri@stonybrook.edu

Names of people whom you discussed the homework with:

1) Naive Bayes with continuous and boolean variables.

$X \to Y$ where $Y$ is boolean

$X$ is $(X_1, X_2)$. $X_1$ is boolean.

$X_2$ is a continuous variable

$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}.$$

Since $Y$ is boolean, $P(Y=1|X) = 1 - P(Y=0|X)$. —— ①

If $P(Y=1/X)$ can be found, $P(Y=0|X)$ can be estimated using ①

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

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

Here $X$ is $(X_1, X_2)$. Using Naive Bayes assumption of independence between $X_1$ and $X_2$. I'm restructuring above term

$$= \frac{P(X_1|Y=1)\,P(X_2|Y=1)\,P(Y=1)}{P(X_1|Y=1)\,P(X_2|Y=1)\,P(Y=1) + P(X_1|Y=0)\,P(X_2|Y=0)\,P(Y=0)}$$

Since $X_1, Y$, taking the boolean variables, I can assume that the distribution followed for these variables is Bernouli.

i.e. $P(X_1|Y=1) = P_1^{X_1}(1-P_1)^{1-X_1}$

$P(X_1|Y=0) = P_2^{X_1}(1-P_2)^{1-X_1}$

$P(Y=0) = P_3$

$P(Y=1) = 1-P_3$.

$$P(x_2|y=k) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \exp\left(\frac{-(x_2-\mu_k)^2}{2\sigma_k^2}\right)$$

(Based on confirmation of piazza post, I'm assuming it as a gaussian distribution).

where $k = 0, 1$

So, all the above equations and the parameters for the distributions are required for $P(y=1|x)$.

$$P(y=0|x) = 1 - P(y=1|x).$$

1.2).

$$P(y=1|x) = \frac{P(x|y=1)\,P(y=1)}{P(x|y=1)\,P(y=1) + P(x|y=0)\,P(y=0)}$$

$$= \frac{1}{1 + \frac{P(x|y=0)\,P(y=0)}{P(x|y=1)\,P(y=1)}} \quad \text{— (2)}$$

$x$ is $(x_1 \ldots x_d)$ vector of boolean variables. I can use naive bayes assumption of independence for these variables.

$$P(x|y=0) = P(x_1|y=0) \ldots P(x_d|y=0) = \prod_{i=1}^{d} P(x_i|y=0) \quad \text{— (3)}$$

Similarly $$P(x|y=1) = \prod_{i=1}^{d} P(x_i|y=1) \quad \text{— (4)}$$

$$P(y=1) = 1 - P(y=0).$$

$$P(y=1) = y$$

$$P(y=0) = 1-y$$

Since $y$ is boolean, I'm using bernouli distribution for this.

From ②,

$$P(y=1|x) = \frac{1}{1 + \exp\left(\log\left(\frac{P(x|y=0)P(y=0)}{P(x|y=1)\,P(y=1)}\right)\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{P(y=0)}{P(y=1)}\right) + \log\left(\frac{P(x|y=0)}{P(x|y=1)}\right)\right)}$$

Using ③ and ④.

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-Y}{Y}\right) + \log\left(\frac{\prod_{i=1}^{d} P(x_i|y=0)}{\prod_{i=1}^{d} P(x_i|y=1)}\right)\right)}$$

Since each $x_i$ is boolean, I can take the bernouli distribution in the below form.

$$P(x_i|y=0) = P_{i0}^{x_i}(1-P_{i0})^{(1-x_i)}$$

$$P(x_i|y=1) = P_{i1}^{x_i}(1-P_{i1})^{1-x_i}$$

This equation satisfies for both 1 and 0 of $x_i$.

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-y}{y}\right) + \sum_{i=1}^{d} \log P(x_i|y=0) - \sum_{i=1}^{d} \log(P(x_i|y=1))\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-y}{y}\right) + \sum_{i=1}^{d}\Big[\big(x_i \log(P_{i0}) + (1-x_i)\log(1-P_{i0})\big) - \big(x_i(\log(P_{i1}) + (1-x_i)\log(1-P_{i1})\big)\Big]\right)}$$

$$= \frac{1}{1 + \exp\left(\log\left(\frac{1-y}{y}\right) + \sum_{i=1}^{d} \log\left(\frac{1-P_{i0}}{1-P_{i1}}\right) + \sum_{i=1}^{d} x_i \left(\log\left(\frac{P_{i0}}{P_{i1}}\right) - \log\left(\frac{1-P_{i0}}{1-P_{i1}}\right)\right)\right)}$$

$$z = \log\left(\frac{1-y}{y}\right) + \sum_{i=1}^{d} \log\left(\frac{1-P_{i0}}{1-P_{i1}}\right)$$

$$K_i = \log\left(\frac{P_{i0}}{P_{i1}}\right) - \log\left(\frac{1-P_{i0}}{1-P_{i1}}\right).$$

$$= \frac{1}{1 + \exp\left(z + \sum_{i=1}^{d} x_i y_i\right)}$$

Hence $P(y=1|x)$ is in logistic regression form.

2.1). Implementation of logistic regression :-

$$P(y=1|\bar{x};\theta) = \frac{1}{1+\exp(-\theta^T\bar{x})} \quad -① \qquad P(y=0|\bar{x};\theta) = \left[\frac{\exp^{-\theta^T\bar{x}}}{1+\exp^{-\theta^T\bar{x}}}\right] \quad -②$$

Applying log and taking differentiation w.r.t $\theta$

$$\log(P(y=1|\bar{x};\theta)) = \log 1 - \log\left(1+\exp^{-(\theta^T\bar{x})}\right)$$

$$\frac{\partial}{\partial\theta} \log(P(y=1|\bar{x};\theta)) = 0 - \frac{1}{\left(1+\exp^{-\theta^T\bar{x}}\right)} \left(\exp^{-\theta^T\bar{x}}\right)(-\bar{x}).$$

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

Similarly for ②,

$$\log\left[P(y=0|\bar{x};\theta)\right] = -\theta^T\bar{x} - \log\left(1+\exp^{-\theta^T\bar{x}}\right)$$

$$\frac{\partial}{\partial\theta}\log\left(P(y=0\mid\bar{x};\theta)\right) = \frac{\partial\left(-\theta^T\bar{x} - \log\left(1+\exp^{-\theta^T\bar{x}}\right)\right)}{\partial\theta}$$

$$= -\bar{x} + \bar{x}\left(\frac{\exp^{-\theta^T\bar{x}}}{1+\exp^{-\theta^T\bar{x}}}\right).$$

---

Combining these two for calculating the $\frac{\partial}{\partial\theta}\log\left(P\left(y^i\mid\bar{x}^i;\theta\right)\right)$

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

$$= y^i\left(\frac{\bar{x}^i\exp^{-\theta^T\bar{x}^i}}{\underbrace{1+\exp^{-\theta^T\bar{x}^i}}_{2}}\right) + (1-y^i)\left(-\bar{x}^i + \bar{x}^i\left(\frac{\exp^{-\theta^T\bar{x}^i}}{\underbrace{1+\exp^{-\theta^T\bar{x}^i}}_{1}}\right)\right)$$

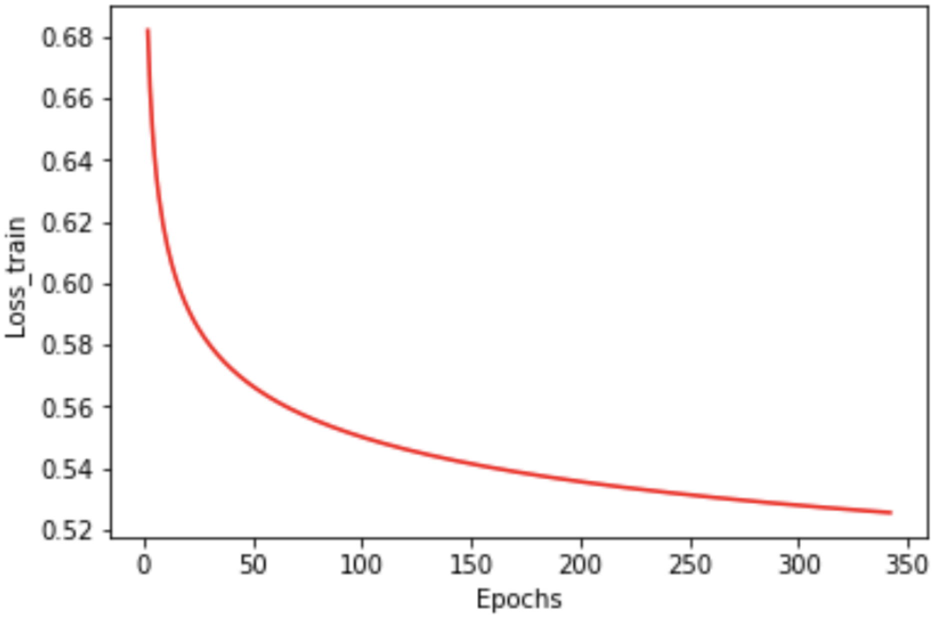
$$= -(1-y^i)\bar{x}^i + \bar{x}^i\left(\frac{\exp^{-\theta^T\bar{x}^i}}{1+\exp^{-\theta^T\bar{x}^i}}\right).$$

1 and 2 terms will be cancelled

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

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

2.3.1)
Please find the answers for the requested questions below:

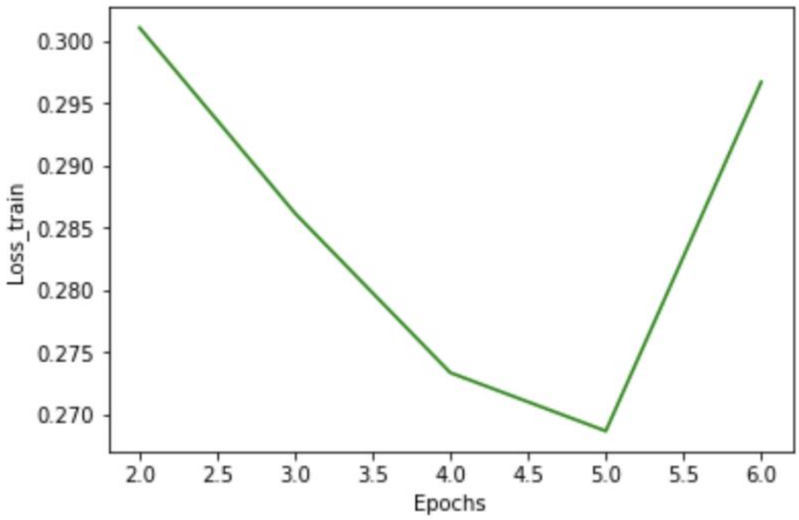



(c) Final value of $L(\theta)$ after optimization = 0.52885375__

Number of epochs till termination = 380
Final value of L(theta)=0.52885375
2.3.2)

Best value for, η0 = _100, η1 = *10*

Number of epochs for training = 5
Final value of L(θ) = 0.2706866

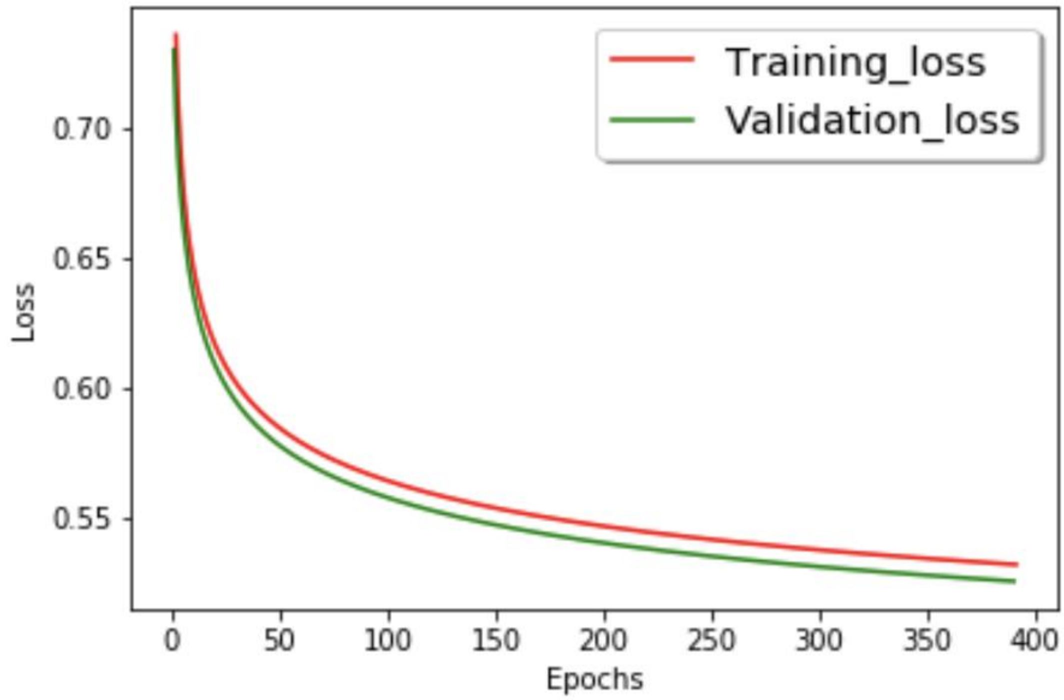


(a) Best value for, $\eta_0$ = _100, $\eta_1$ = \10

Number of epochs for training = 5__

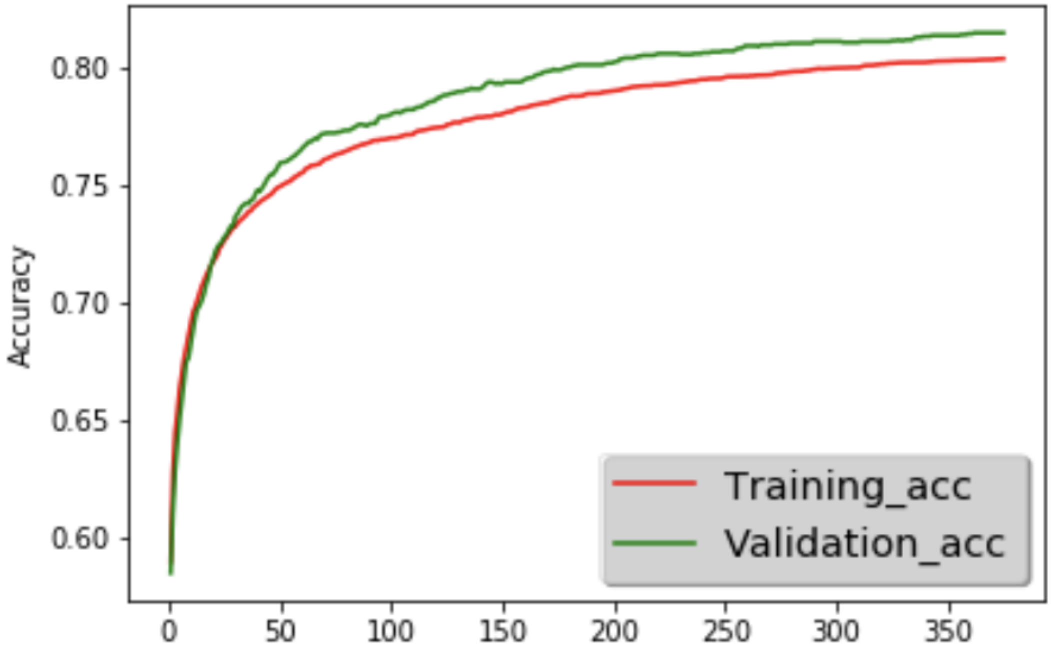
Final value of L($\theta$) = 0.2706866_

2.3.3)



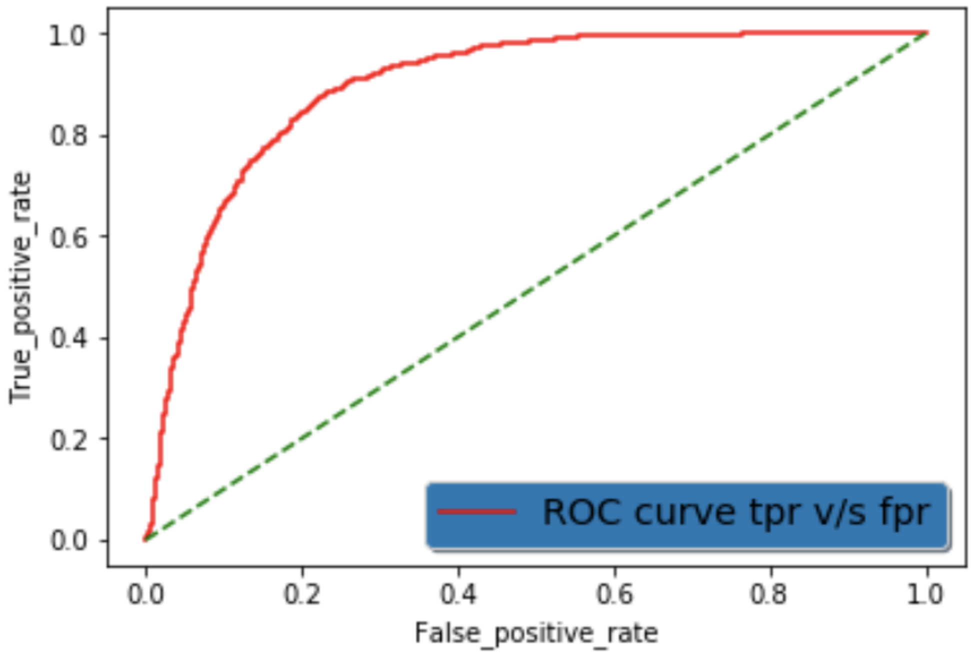

Please find the above plot for the training loss.

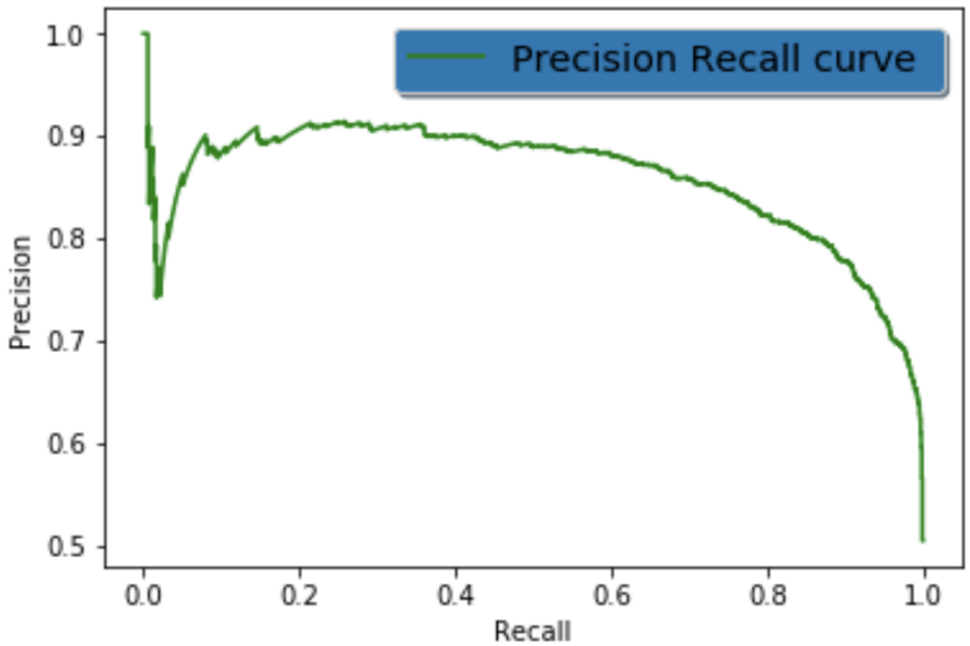
2.3.3)




Please find the above plot for the training accuracy.

2.3.4)




```
Area Under Curve for fpr and tpr 0.8955787571615552
```

Please find the above plot for the ROC and precision recall curve.
Average under the curve for tpr and fpr is 0.89557161

Average precision curve 0.8586262157642053

My Kaggle score is
Best obtained accuracy on Public Leader-board = 0.881276