# CIS 519: Homework 4

{Ian MacDonald}

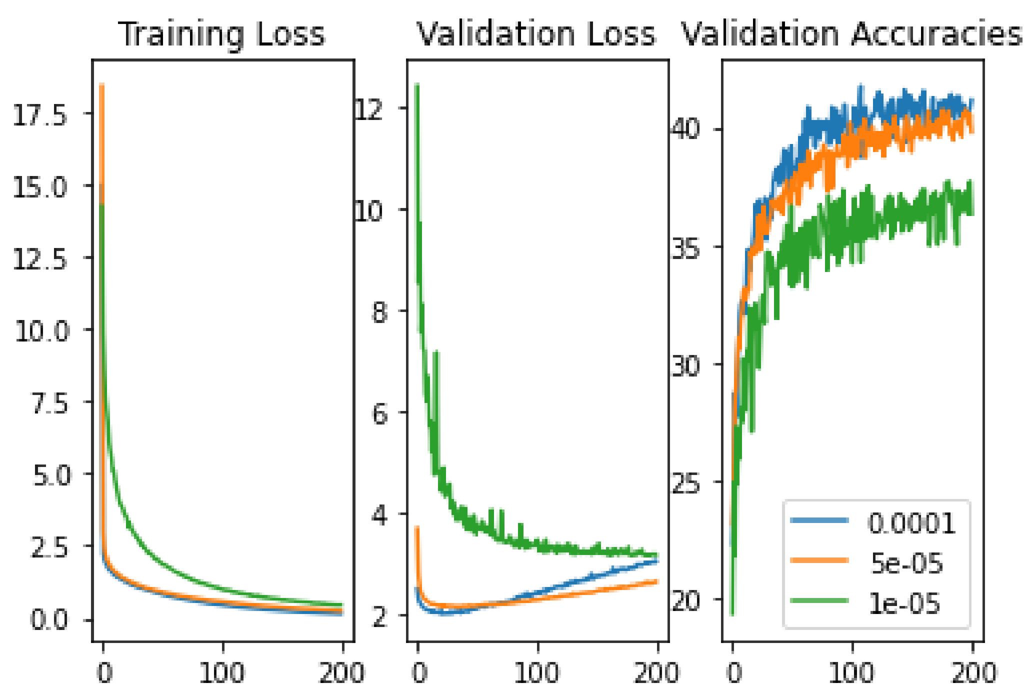
12/07/20

Although the solutions are my own, I consulted with the following people while working on this homework: {Sally Hu}

## Neural Networks

### Feed Forward

Plots:

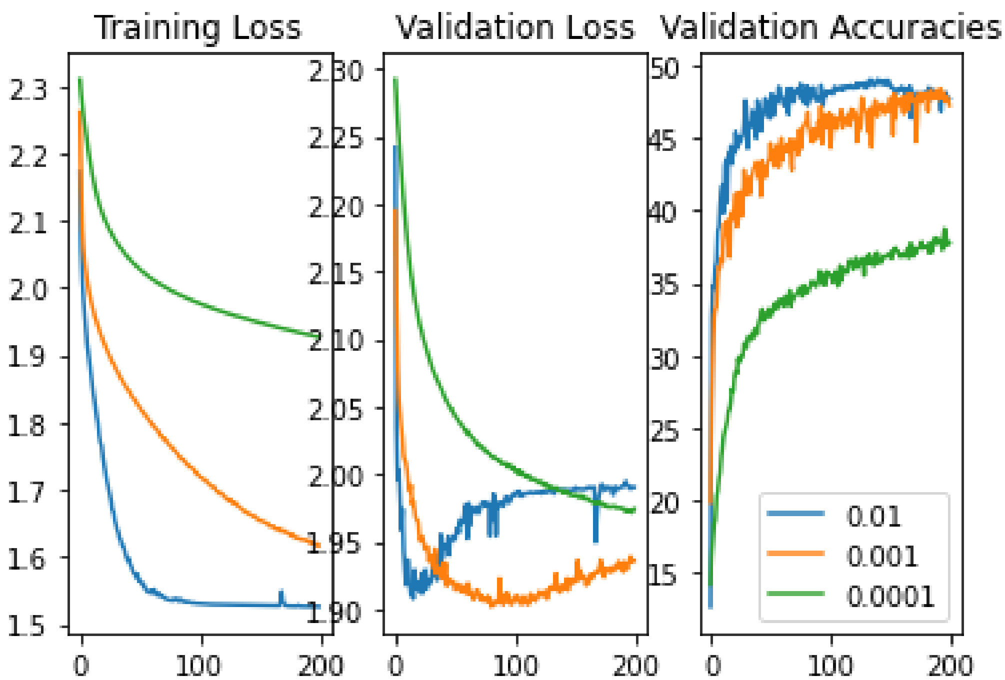


| Learning Rate | Final Validation Accuracy |
|---|---|
| 0.0001 | 41.15 |
| 0.00005 | 39.8 |
| 0.00001 | 36.3 |

Learning rate with best validation accuracy: 0.0001 The test accuracy of that model is 40.2

## Convolutional

Plots:




| Learning Rate | Final Validation Accuracy |
|---|---|
| 0.01 | 47.7 |
| 0.001 | 47.25 |
| 0.0001 | 37.8 |

Learning rate with best validation accuracy: 0.01 The test accuracy of that model is 47.9

# Document Classification

| Representation | $k$ | Test Accuracy |
|---|---|---|
| BBoW | 0.1 | .86 |
| CBoW | 0.1 | .8608 |
| TF-IDF | 1.0 | .83 |

What impact does $k$ have? That is, as the value of $k$ goes to infinity, what will happen to $P(y \mid d)$?

k is a smoothing factor, which gives previously unseen values a low probability to occur in either class, rather than a probability of 0, allowing us to better handle documents where some tokens do not appear in the vocabulary. If the value of k goes to infinity, then $P(y \mid d)$ will converge to $\frac{1}{|V|}$, which is a uniform distribution where every element in the vocabulary has an equal chance to be a part of the document regardless of the document's classification. This makes sense intuitively, as increasing k increases the smoothness, and infinite smoothness would just be a uniform distribution.

# Theory

## Multivariate Exponential Naive Bayes

1.

| | |
|---|---|
| $P(Y{=}A) = .429$ | $P(Y{=}B) = .571$ |
| $\lambda_{A;1} = .375$ | $\lambda_{B;1} = .222$ |
| $\lambda_{A;2} = .158$ | $\lambda_{B;2} = .286$ |

2. $\frac{e^{\frac{-9}{8}} * e^{\frac{-15}{19}} * \frac{3}{8} * \frac{3}{19}}{e^{\frac{-12}{18}} * e^{\frac{-20}{14}} * \frac{4}{18} * \frac{4}{14}}$

3. if $\frac{e^{\frac{-9}{8}} * e^{\frac{-15}{19}} * \frac{3}{8} * \frac{3}{19}}{e^{\frac{-12}{18}} * e^{\frac{-20}{14}} * \frac{4}{18} * \frac{4}{14}} * \frac{P(Y=A)}{P(Y=B)} > 1$, Predict Y=A

4. Classifier predicts Y=B

Work for above question:
Given p as $P(Y{=}A)$
L(p) = $p^3(1-p)^4$
L'(p) = $(1-p)^4 3p^2 - 4p^3(1-p)^3$
$0 = (1-p)^4 3p^2 - 4p^3(1-p)^3$
$0 = p^2(1-p)^3(3-7p)$
MLE estimation for $P(Y{=}A) = \frac{3}{7}$
MLE estimation for $P(Y{=}B) = 1 - P(Y{=}A) = \frac{4}{7}$

L($\lambda_{A;1}$) = $e^{-\lambda}\lambda * e^{-3\lambda}\lambda * e^{-4\lambda}\lambda$
L($\lambda_{A;1}$) = $\lambda^3 e^{-8\lambda}$

Log L($\lambda_{A;1}$) = $\log(\lambda^3 e^{-8\lambda}$
Log L($\lambda_{A;1}$) = $3ln(\lambda) + -8\lambda ln(e)$
Log L'($\lambda_{A;1}$) = $\frac{3}{\lambda} - 8$
$0 = \frac{3}{\lambda} - 8$
$\lambda_{A;1} = \frac{3}{8}$
MLE estimation for $\lambda_{A;1}$ is .375

L($\lambda_{A;2}$) = $e^{-4\lambda}\lambda * e^{-9\lambda}\lambda * e^{-6\lambda}\lambda$
L($\lambda_{A;2}$) = $\lambda^3 e^{-19\lambda}$
Log L($\lambda_{A;2}$) = $\log(\lambda^3 e^{-8\lambda}$
Log L($\lambda_{A;2}$) = $3ln(\lambda) + -19\lambda ln(e)$
Log L'($\lambda_{A;2}$) = $\frac{3}{\lambda} - 19$
$0 = \frac{3}{\lambda} - 19$
$\lambda_{A;2} = \frac{3}{19}$
MLE estimation for $\lambda_{A;2}$ is .159

L($\lambda_{B;1}$) = $e^{-7\lambda}\lambda * e^{-2\lambda}\lambda * e^{-3\lambda}\lambda * e^{-6\lambda}\lambda$
L($\lambda_{B;1}$) = $\lambda^4 e^{-18\lambda}$
Log L($\lambda_{B;1}$) = $\log(\lambda^4 e^{-18\lambda}$
Log L($\lambda_{B;1}$) = $4ln(\lambda) + -18\lambda ln(e)$
Log L'($\lambda_{B;1}$) = $\frac{4}{\lambda} - 18$
$0 = \frac{4}{\lambda} - 18$
$\lambda_{B;1} = \frac{4}{18}$
MLE estimation for $\lambda_{B;1}$ is .222

L($\lambda_{B;2}$) = $e^{-3\lambda}\lambda * e^{-6\lambda}\lambda * e^{0\lambda}\lambda * e^{-5\lambda}\lambda$
L($\lambda_{B;2}$) = $\lambda^4 e^{-14\lambda}$
Log L($\lambda_{B;2}$) = $\log(\lambda^4 e^{-14\lambda}$
Log L($\lambda_{B;2}$) = $4ln(\lambda) + -14\lambda ln(e)$
Log L'($\lambda_{B;2}$) = $\frac{4}{\lambda} - 14$
$0 = \frac{4}{\lambda} - 14$
$\lambda_{B;2} = \frac{4}{14}$
MLE estimation for $\lambda_{B;2}$ is .286

Logic for part 3:
$P(Y{=}A|X) = \frac{P(X|Y=A)*P(Y=A)}{P(X)}$
$P(Y{=}B|X) = \frac{P(X|Y=B)*P(Y=B)}{P(X)}$
$\frac{P(Y=A|X)}{P(Y=B|X)} = \frac{P(X|Y=A)*P(Y=A)}{P(X|Y=B)*P(Y=B)}$

## Coin Toss

The most likely value of $p$ is .632

Explanation:
Probability of H=$p$, probability of T=$1 - p$

If flip is T, it is shown as is. If it is H, it is flipped again and then shown. Therefore, sequence to get a T can be T or HT, while to get an H is must be HH. Actual probability of seeing a H shown is $p^2$, and $1 - p^2$ for T.

$L(p) = (1 - p^2)^6 (p^2)^4$
$L(p) = (1 - p^2)^6 (p^8)$
$L'(p) = (1 - p^2)^6 (8p^7) + (p^8) - 12p(1 - p^2)^5$
$L'(p) = (1 - p^2)^6 (8p^7) - 12p^9 (1 - p^2)^5$
$L'(p) = (1 - p^2)^5 (p^7)(8 - 20p^2)$
$L'(p) = (1 - p^2)^5 (p^7)(2 - 5p^2)$
$0 = (2 - 5p^2)$
$p = .632$