COMS 4771 HW 1

Griffin Klett

TOTAL POINTS

82 / 115

QUESTION 1

1 Analyzing Iterative Optimization 6 / 30

✓ **+ 6 pts (i) Correct**

+ **4 pts** (i) Did not prove symmetric
+ **2 pts** (i) Did not prove positive semi-definite
- **1 pts** (i) Minor error
+ **0 pts** (i) Incorrect or missing
+ **8 pts** (ii) Correct
+ **7 pts** (ii) Small mistake
+ **5 pts** (ii) Medium Mistake
+ **3 pts** (ii) Big mistake/Insufficient proof

✓ **+ 0 pts (ii) Completely incorrect or missing**

+ **8 pts** (iii) Correct
+ **7 pts** (iii) Small mistake
+ **6 pts** (iii) Medium Mistake
+ **4 pts** (iii) Large mistake

✓ **+ 0 pts (iii) Incorrect or missing (no credit if proof not attempted)**

+ **8 pts** (iv) Correct

✓ **+ 0 pts (iv) Incorrect or missing**

+ **0 pts** Plagiarism (see comments for exact deduction)
- **1 pts** Missing reference to the fact that new matrix is PSD and symmetric
- **3 pts** Major error in final steps
- **2 pts** Error in eigenvalue computation
- **1 pts** Notational issues
- **1 pts** Minor details missing (eg. unjustified use of identities,etc)
- **2 pts** Error in Cauchy Schwartz/early computation
- **2 pts** Minor error in final steps

QUESTION 2

2 Statistical Estimators 20 / 25

✓ **- 0 pts (i) Correct**

- **3 pts** (i) distribution derived incorrectly
- **4 pts** (i) MLEs are incorrect
- **7 pts** (i) Incorrect/Missing
- **0 pts** (ii) Correct
- **3 pts** (ii) Minor Issue/ Didn't prove the case when g is not one-to-one

✓ **- 5 pts (ii) Proof is insufficient**

- **10 pts** (ii) Incorrect/Missing

✓ **- 0 pts (iii) Correct**

- **2 pts** (iii) Consistent and Unbiased - 2 incorrect examples
- **1 pts** (iii) Consistent and Unbiased - 1 incorrect example
- **2 pts** (iii) Consistent and Biased - 2 incorrect examples
- **1 pts** (iii) Consistent and Biased - 1 incorrect example
- **2 pts** (iii) Inconsistent and Unbiased - 2 incorrect examples
- **1 pts** (iii) Inconsistent and Unbiased - 1 incorrect example
- **2 pts** (iii) Inconsistent and Biased - 2 incorrect examples
- **1 pts** (iii) Inconsistent and Biased - 1 incorrect example
- **8 pts** (iii) Incorrect/Missing Solution or Proof

QUESTION 3

3 Evaluating Classifiers 16 / 20

+ **2 pts** i) Partial Credit

✓ **+ 4 pts i) Full Credit**

+ **2 pts** ii) Partial Credit - Explain why a=b without deriving it
+ **4 pts** ii) Partial Credit - Attempt analysis of expected error but fail at identifying a=b or identifies a=b with substantial errors

+ **6 pts** ii) Partial Credit - Identify a=b, with small errors in derivation

✓ + **8 pts** **ii) Full Credit**

+ **2 pts** iii) Partial Credit - attempted

+ **6 pts** iii) Partial Credit - mostly correct with small error(s)

✓ + **4 pts** **iii) Partial Credit - somewhat correct but with significant error(s)**

+ **8 pts** iii) Full Credit

+ **0 pts** No work

+ **0 pts** Handwritten

1 y1 and y2 should be reversed

QUESTION 4

4 Email Spam Classification Case Study **40 / 40**

✓ - **0 pts** **No major errors**

- **5 pts** partially missing classifier or mistake
- **10 pts** missing an entire classifier or wrong implementation
- **20 pts** Code not present or major part missing in code
- **25 pts** Incorrect or not present 4 ii)
- **5 pts** Incorrect/invalid/missing justification for 4 iii)
- **10 pts** Missing performance graphs for 4 iii)
- **2 pts** Missing justified final decision on best model for 4 iii)
- **2.5 pts** Partially incorrect justification for 4 iii)
- **5 pts** Partially missing performance analysis or graphs
- **18 pts** Code Copied From External Sources Adjustment

next time when submitting file please make sure you are not submitting the IDE's config files

QUESTION 5

5 Adjustments **0 / 0**

✓ - **0 pts** **No adjustments**

- **3 pts** Pages assigned incorrectly
- **3 pts** Code submitted incorrecltly
- **115 pts** HOLD (please contact the instructor directly to release the hold)

COMS 4771 Machine Learning (Spring 2021)
Problem Set #1

Griffin Klett - Gk2591@columbia.edu

2/13/2021

Problem 1: Analyzing Iterative Optimization

(i)
Definition: For positive semidefinite matrix, quadratic form is ≥ 0; Quadratic form of M is $\vec{x} \cdot M\vec{x}$

$$Q(x) = \vec{x} \cdot M\vec{x} = \vec{x} \cdot A^T A\vec{x} \tag{1}$$

$$Q(x) = \vec{x}^T A^T A\vec{x} \tag{2}$$

$$Q(x) = (A\vec{x})^T A\vec{x} \tag{3}$$

$$Q(x) = A\vec{x} \cdot A\vec{x} \tag{4}$$

$$Q(x) = ||A\vec{x}||^2 \tag{5}$$

$$Q(x) = ||A\vec{x}||^2 \text{ which is always } \geq 0 \tag{6}$$

Symmetric definition:

$$M = A^T A \tag{7}$$

$$M^T = (A^T A)^T = A^T A \tag{8}$$

(ii)
(iii)
(iv)

1 Analyzing Iterative Optimization 6 / 30

- ✓ **+ 6 pts** **(i) Correct**
- **+ 4 pts** (i) Did not prove symmetric
- **+ 2 pts** (i) Did not prove positive semi-definite
- **- 1 pts** (i) Minor error
- **+ 0 pts** (i) Incorrect or missing
- **+ 8 pts** (ii) Correct
- **+ 7 pts** (ii) Small mistake
- **+ 5 pts** (ii) Medium Mistake
- **+ 3 pts** (ii) Big mistake/Insufficient proof
- ✓ **+ 0 pts** **(ii) Completely incorrect or missing**
- **+ 8 pts** (iii) Correct
- **+ 7 pts** (iii) Small mistake
- **+ 6 pts** (iii) Medium Mistake
- **+ 4 pts** (iii) Large mistake
- ✓ **+ 0 pts** **(iii) Incorrect or missing (no credit if proof not attempted)**
- **+ 8 pts** (iv) Correct
- ✓ **+ 0 pts** **(iv) Incorrect or missing**
- **+ 0 pts** Plagiarism (see comments for exact deduction)
- **- 1 pts** Missing reference to the fact that new matrix is PSD and symmetric
- **- 3 pts** Major error in final steps
- **- 2 pts** Error in eigenvalue computation
- **- 1 pts** Notational issues
- **- 1 pts** Minor details missing (eg. unjustified use of identities,etc)
- **- 2 pts** Error in Cauchy Schwartz/early computation
- **- 2 pts** Minor error in final steps

Problem 2 Statistical Estimators

(i)

$$MLE(\theta) = Pr(X|\theta) = pr(x_1) \cdot pr(x_2) \cdot \ldots \cdot pr(x_n) \tag{9}$$

In order to maximize the probability of seeing the samples $x_1 \ldots x_n$, we want $p(x_i|\theta) = 1$, which happens in the case where $a \leq x \leq b$. After seeing the samples, we can set $a = min(x_1 \ldots x_n)$ and $b = max(x_1 \ldots x_n)$. Then we can calculate the likelihood:

$$Pr(X|\theta) = pr(x_1) \cdot pr(x_2) \cdot \ldots \cdot pr(x_n) = 1 \cdot 1 \cdot \ldots \cdot 1 = MLE(\theta) = 1 \tag{10}$$

The maximum likelihood estimate of 1 is the highest possible, so the parameters for a and b are optimal.

$$\text{Answer: } a = min(x_1 \ldots x_n) \text{ , } b = max(x_1 \ldots x_n) \tag{11}$$

(ii)
show that the MLE of $\theta = g(\theta_{ML}$

$$\text{Since} f_n[x|\theta] \text{ is maximized when } \theta = \hat{\theta} \tag{12}$$

$$f_n[x|h(\psi)] \text{ is maximized when } h(\psi) = \hat{\theta} \tag{13}$$

$$\text{So } h(\hat{\psi}) = \hat{\theta} \text{ and } \hat{\psi} = g(\hat{\theta}) \tag{14}$$

(iii)
Inconsistent, biased, Example 1:

$$\hat{\mu} = X_1 + X_2 \tag{15}$$

Bias shown:

$$E[\hat{\mu}] = 2\mu \neq \mu \tag{16}$$

Inconsistency shown:

$$n \to \infty, Var(X_1 + X_2) = \underbrace{Var(X_1) + Var(X_2)}_{\text{Constant; does not approach value of estimator nor variance approach 0}} \tag{17}$$

Inconsistent, biased, Example 2:

$$\hat{\mu} = \frac{X_1 + X_2}{3} \tag{18}$$

Bias shown:

$$E[\hat{\mu}] = \frac{2\mu}{3} \neq \mu \tag{19}$$

Inconsistency shown:

$$n \to \infty, Var(\frac{1}{3}X_1 + \frac{1}{3}X_2) = \underbrace{\frac{1}{9}Var(X_1) + \frac{1}{9}Var(X_2)}_{\text{Constant; does not approach value of estimator and variance does not approach 0}} \tag{20}$$

Inconsistent, Unbiased, Example 1:

$$\hat{\mu} = X_1 \tag{21}$$

Unbiased shown:

$$E[\hat{\mu}] = \mu \tag{22}$$

Inconsistency shown:

$$n \to \infty, Var(X_1) = \underbrace{Var(X_1) = \sigma^2}_{\text{Does not approach value of estimator nor variance approach 0}} \tag{23}$$

Inconsistent, unbiased, Example 2:

$$\hat{\mu} = \frac{X_1 + X_2}{2} \tag{24}$$

unbiased shown:

$$E[\hat{\mu}] = \frac{2\mu}{2} = \mu \tag{25}$$

Inconsistency shown:

$$n \to \infty, Var(X_1 + X_2) = \underbrace{\frac{1}{2}Var(X_1) + \frac{1}{2}Var(X_2)}_{\text{Constant; does not approach value of estimator and variance does not approach 0}} \tag{26}$$

Consistent, biased, Example 1:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i) + \frac{1}{n^2} \tag{27}$$

Biased shown:

$$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}(E[x_i]) + \frac{1}{n^2} = \mu + \frac{1}{n^2} \neq \mu \tag{28}$$

Consistency shown:

$$E[\hat{\mu}] = lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(E[x_i]) + \frac{1}{n^2} = \frac{n\mu}{n} + 0 = \mu \tag{29}$$

Consistent, biased, Example 2:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i) + \frac{1}{n} \tag{30}$$

Biased shown:

$$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}(E[x_i]) + \frac{1}{n} = \mu + \frac{1}{n} \neq \mu \tag{31}$$

Consistency shown:

$$E[\hat{\mu}] = lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(E[x_i]) + \frac{1}{n} = \frac{n\mu}{n} + 0 = \mu \tag{32}$$

Consistent, unbiased, Example 1:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{33}$$

unbiased shown:

$$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}(E[x_i]) = \mu \tag{34}$$

Consistency shown:

$$E[\hat{\mu}] = lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(E[x_i]) = \frac{n\mu}{n} = \mu \tag{35}$$

Consistent, unbiased, Example 2:

$$\hat{\mu} = \frac{1}{n-1}\sum_{i=2}^{n}(x_i) \tag{36}$$

unbiased shown:

$$E[\hat{\mu}] = \frac{1}{n-1}\sum_{i=2}^{n}(E[x_i]) = \frac{\mu(n-1)}{n-1} = \mu \tag{37}$$

Consistency shown:

$$E[\hat{\mu}] = lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}(E[x_i]) = \frac{n\mu}{n} = \mu \tag{38}$$

2 Statistical Estimators **20 / 25**

- ✓ **- 0 pts** **(i) Correct**
- **- 3 pts** (i) distribution derived incorrectly
- **- 4 pts** (i) MLEs are incorrect
- **- 7 pts** (i) Incorrect/Missing
- **- 0 pts** (ii) Correct
- **- 3 pts** (ii) Minor Issue/ Didn't prove the case when g is not one-to-one
- ✓ **- 5 pts** **(ii) Proof is insufficient**
- **- 10 pts** (ii) Incorrect/Missing
- ✓ **- 0 pts** **(iii) Correct**
- **- 2 pts** (iii) Consistent and Unbiased - 2 incorrect examples
- **- 1 pts** (iii) Consistent and Unbiased - 1 incorrect example
- **- 2 pts** (iii) Consistent and Biased - 2 incorrect examples
- **- 1 pts** (iii) Consistent and Biased - 1 incorrect example
- **- 2 pts** (iii) Inconsistent and Unbiased - 2 incorrect examples
- **- 1 pts** (iii) Inconsistent and Unbiased - 1 incorrect example
- **- 2 pts** (iii) Inconsistent and Biased - 2 incorrect examples
- **- 1 pts** (iii) Inconsistent and Biased - 1 incorrect example
- **- 8 pts** (iii) Incorrect/Missing Solution or Proof

Problem 3: Evaluating Classifiers

(i)

$$P[f_t(x) \neq y] = P[f_t(x) = y_1 \text{ and } y_{true} = y_2] + P[f_t(x) = y_2 \text{ and } y_{true} = y_1] \tag{39}$$

We know further that these two above cases happen when x ¿ t and $y_{true} = y_2$ or $x \leq t$ and $y_{true} = y_1$.

$$P[f_t(x) \neq y] = P[x > t \text{ and } y_{true} = y_2] + P[x \leq t \text{ and } y_{true} = y_1] \tag{40}$$

integrating with respect to x

$$P[f_t(x) \neq y] = \int_t^{\infty} P(X = x, y = y_2)\, dx + \int_{-\infty}^{t} P(X = x, y = y_1)\, dx \tag{41}$$

(ii) For optimal value, take derivative and set equal to zero

$$\frac{d}{dt}P[f_t(x) \neq y] = \frac{d}{dt}(\int_t^{\infty} P(X = x, y = y_2)\, dx) + \frac{d}{dt}(\int_{-\infty}^{t} P(X = x, y = y_1)\, dx) = 0 \tag{42}$$

flip the bounds of the second integral (negative integral from t -¿ $-\infty$, and evaluate at end points, cancelling terms, you get:

$$\frac{d}{dt}P[f_t(x) \neq y] = P(X = t, y = y_2) - P(X = t, y = y_1) = 0 \tag{43}$$

$$P(X = t, y = y_2) = P(X = t, y = y_1) \tag{44}$$

And apply Bayes theorem:

$$P(X = t|Y = y_2) \cdot P(Y = y_2) = P(X = t|Y = y_1) \cdot P(Y = y_2) \tag{45}$$

(iii) since the class priors are equal, we can rewrite the equation from part two as:

$$P(X = t|Y = y_2) = P(X = t|Y = y_1) \tag{46}$$

and our classifier makes the choice about y_1 or y_2 based only on the class conditionals:

$$\hat{f}(x) = argmax_{y \epsilon Y} P(X = x|Y = y) \tag{47}$$

To achieve Bayes error rate then we just need to have the means for the distributions of y_1's to the left of the mean of the distribution of y_2's. Then there will be some threshold t, between these two, where we switch from predicting y_1 toy_2.

Answer:
$P(X = x \parallel Y = y_1) \sim N(10, 5)$
P(X = x $\parallel Y = y_2) \sim N(15, 5)$

In this case, the threshold value t would be somewhere between 10 and 15.

In the opposite case, if the mean of the distribution for y_1 is to the right of the mean of the distribution for y_2, then there will be no such value of t where we could achieve Bayes error rate.

Answer:
$P(X = x \parallel Y = y_1) \sim N(15, 5)$
P(X = x $\parallel Y = y_2) \sim N(10, 5)$

3 Evaluating Classifiers **16 / 20**

+ **2 pts** i) Partial Credit

✓ + **4 pts** **i) Full Credit**

+ **2 pts** ii) Partial Credit - Explain why a=b without deriving it

+ **4 pts** ii) Partial Credit - Attempt analysis of expected error but fail at identifying a=b or identifies a=b with substantial errors

+ **6 pts** ii) Partial Credit - Identify a=b, with small errors in derivation

✓ + **8 pts** **ii) Full Credit**

+ **2 pts** iii) Partial Credit - attempted

+ **6 pts** iii) Partial Credit - mostly correct with small error(s)

✓ + **4 pts** **iii) Partial Credit - somewhat correct but with significant error(s)**

+ **8 pts** iii) Full Credit

+ **0 pts** No work

+ **0 pts** Handwritten

1 y1 and y2 should be reversed

Problem 4: Email Spam Classification Case Study

(i) Bag of words embedding done in python scripts
(ii) Python scripts have the classifiers
(iii) Attached on next page

(iii).

Conclusion:

Best Naïve Bayes Model: 97.48%

Best Decision Tree Model: 91.802%

Best KNN Model: 91.4%

Overall, the Bayes model performed the best after tuning all three types of models. This type of model was robust to smaller sample/training sizes, but still incorporates every "piece"/word of data into the decision. As opposed to decision tree, in which only certain keywords are used to make the decision. Even if most of the data is useful (it is not), only branching on certain features leaves out much of the valuable information from other words. I think therefore that decision trees are not the most appropriate model for dealing with bag of words and spam classifications. Similarly, in the decision tree, I feel it is easier to overfit with the decision tree, as one word can change the classification. KNN way less robust as training size decreased; this is expected, as the proof of KNN approaching Bayes classification relies on the sample size increasing to infinity, as the likelihood of finding another datapoint exactly like the test one increases.

Comparison of each individual model tuning/parameters below.

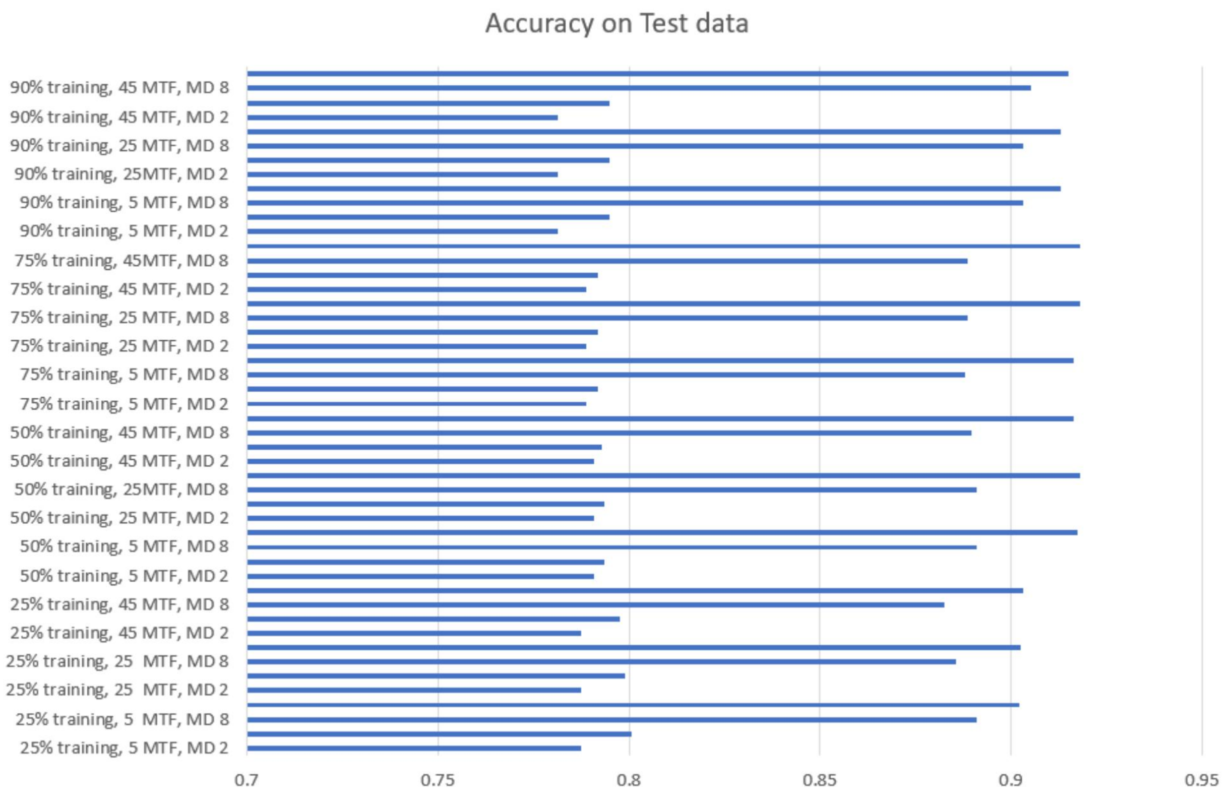
Decision Tree:

In order to get the best Decision Tree classifier, I experimented with three variables: the percent of data split to training/testing, the minimum term frequency (shown in image as MTF), and Max Tree Depth (MD). The justification for experimenting with minimum term frequency is that I believed very infrequent words (appearing less than 5, 25, or 45 times throughout the entire training dataset) would not be predictive and would probably contribute more to noise then the signal. Similarly with max tree depth, I did not want to overfit the training data, which is possible using a large tree depth. I used Gini Impurity and information gain to decide when to branch.

Conclusions from Decision Tree:

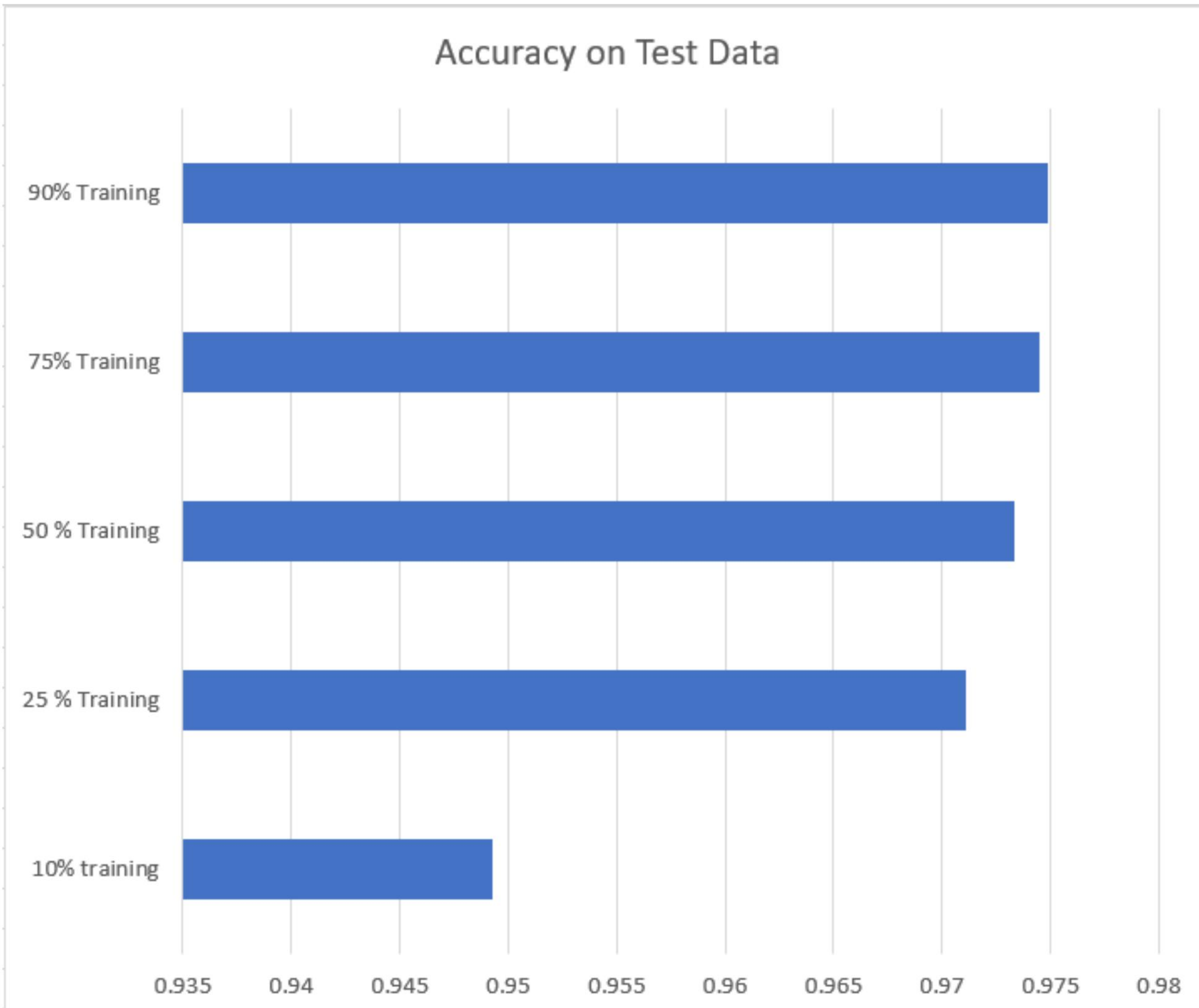
From below and the appendix with the exact figures, the highest accuracy achieved was 91.802% classification; two models had this exact value. Both were using 75% of the dataset to train, had a max depth of 12, and minimum word frequencies of 25 or 45.

Because the higher term frequency requirement makes the classification faster, I have chosen to use this model in my final submission.


Accuracy on Test data
90% training, 45 MTF, MD 8
90% training, 45 MTF, MD 2
90% training, 25 MTF, MD 8
90% training, 25MTF, MD 2
90% training, 5 MTF, MD 8
90% training, 5 MTF, MD 2
75% training, 45MTF, MD 8
75% training, 45 MTF, MD 2
75% training, 25 MTF, MD 8
75% training, 25 MTF, MD 2
75% training, 5 MTF, MD 8
75% training, 5 MTF, MD 2
50% training, 45 MTF, MD 8
50% training, 45 MTF, MD 2
50% training, 25MTF, MD 8
50% training, 25 MTF, MD 2
50% training, 5 MTF, MD 8
50% training, 5 MTF, MD 2
25% training, 45 MTF, MD 8
25% training, 45 MTF, MD 2
25% training, 25 MTF, MD 8
25% training, 25 MTF, MD 2
25% training, 5 MTF, MD 8
25% training, 5 MTF, MD 2
0.7
0.75
0.8
0.85
0.9
0.95

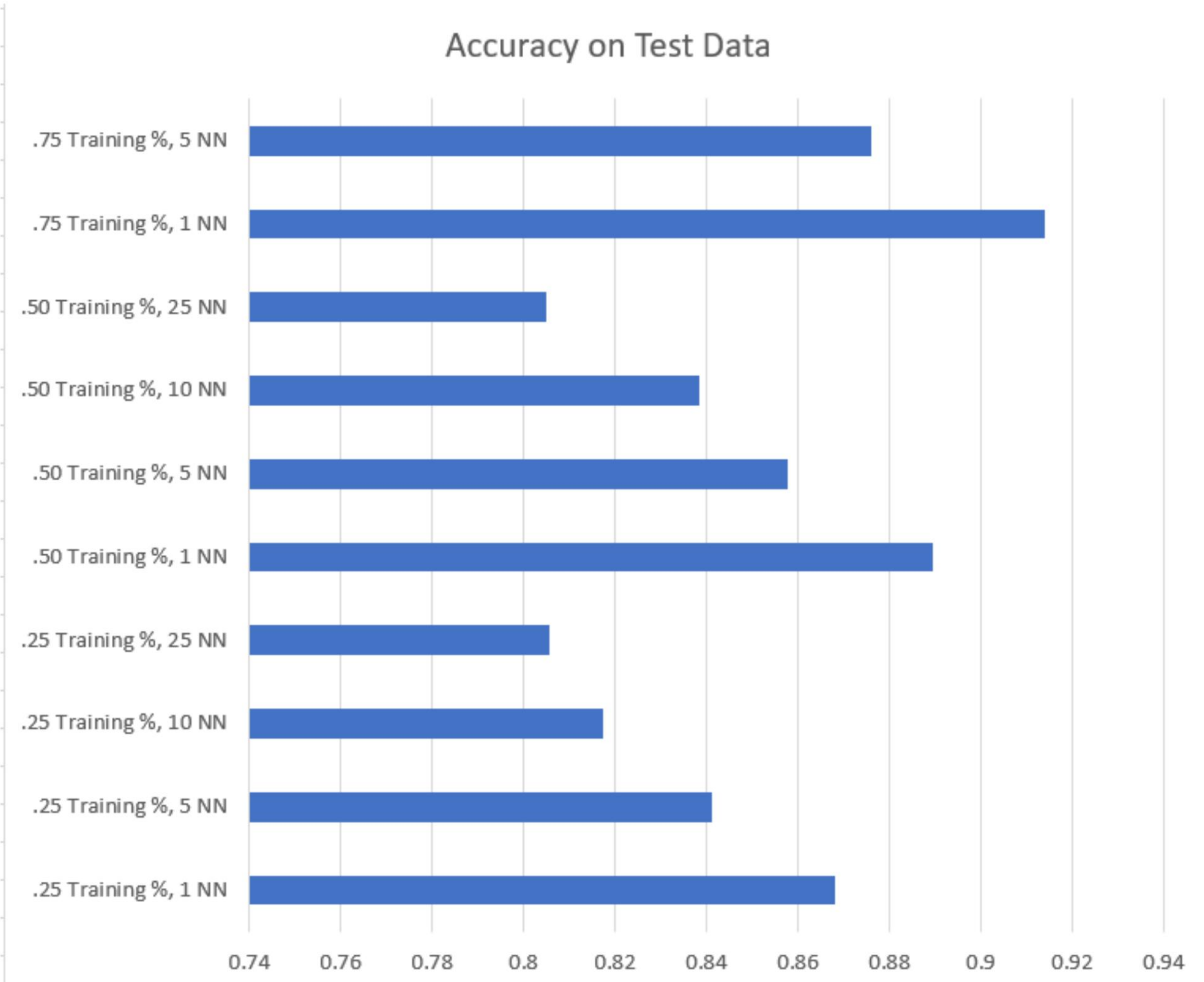
Naïve-Bayes:

With Naïve Bayes, there were no hyperparameters to tune, so the accuracy shows only the different values for different splits.



KNN:

In order to get the best KNN classifier, I experimented with the number of neighbors used for the classification, the training %, and the distance ord (L1, L2, etc), and below is the accuracy for some of the various runs, with the exact values in the appendix below.



Conclusions for KNN:

From this we can see that the 1 NN performed best at all the various splits of training data, and the .75% split of train-test gave the highest accuracy.
The best model uses 1 NN and 75% of train/test data, which yields an accuracy of 91.4% on test data.

Decision Tree Appendix:

Training split	min term frequency	max tree depth	correct ham classification	incorrect ham classififcation	correct spam classification	incorrect spam classification	total correct	total incorrect	accuracy
0.25	5	2	2701	53	353	772	3054	825	0.787316319
0.25	5	4	2036	718	1070	55	3106	773	0.800721836
0.25	5	8	2498	256	958	167	3456	423	0.890951276
0.25	5	12	2498	256	1001	124	3499	380	0.902036607
0.25	25	2	2701	53	353	772	3054	825	0.787316319
0.25	25	4	2030	724	1069	56	3099	780	0.798917247
0.25	25	8	2486	268	949	176	3435	444	0.88553751
0.25	25	12	2505	249	996	129	3501	378	0.902552204
0.25	45	2	2701	53	353	772	3054	825	0.787316319
0.25	45	4	2026	728	1068	57	3094	785	0.797628255
0.25	45	8	2474	280	950	175	3424	455	0.882701727
0.25	45	12	2506	248	997	128	3503	376	0.903067801
0.5	5	2	1800	36	245	505	2045	541	0.790796597

0.5	5	4	1343	493	709	41	2052	534	0.79350348
0.5	5	8	1585	251	719	31	2304	282	0.89095127 6
0.5	5	12	1656	180	716	34	2372	214	0.91724671 3
0.5	25	2	1800	36	245	505	2045	541	0.79079659 7
0.5	25	4	1343	493	709	41	2052	534	0.79350348
0.5	25	8	1579	257	725	25	2304	282	0.89095127 6
0.5	25	12	1652	184	722	28	2374	212	0.91802010 8
0.5	45	2	1800	36	245	505	2045	541	0.79079659 7
0.5	45	4	1342	494	708	42	2050	536	0.79273008 5
0.5	45	8	1581	255	720	30	2301	285	0.88979118 3
0.5	45	12	1653	183	717	33	2370	216	0.91647331 8
0.75	5	2	899	19	121	254	1020	273	0.78886310 9
0.75	5	4	667	251	357	18	1024	269	0.79195669
0.75	5	8	788	130	360	15	1148	145	0.88785769 5
0.75	5	12	820	98	365	10	1185	108	0.91647331 8

0.75	25	2	899	19	121	254	1020	273	0.788863109
0.75	25	4	667	251	357	18	1024	269	0.7919566 9
0.75	25	8	788	130	361	14	1149	144	0.8886310 9
0.75	25	12	821	97	366	9	1187	106	0.918020108
0.75	45	2	899	19	121	254	1020	273	0.788863109
0.75	45	4	667	251	357	18	1024	269	0.7919566 9
0.75	45	8	788	130	361	14	1149	144	0.8886310 9
0.75	45	12	821	97	366	9	1187	106	0.918020108
0.9	5	2	359	8	45	105	404	113	0.781431335
0.9	5	4	267	100	144	6	411	106	0.794970986
0.9	5	8	324	43	143	7	467	50	0.903288201
0.9	5	12	334	33	138	12	472	45	0.912959381
0.9	25	2	359	8	45	105	404	113	0.781431335
0.9	25	4	267	100	144	6	411	106	0.794970986
0.9	25	8	324	43	143	7	467	50	0.903288201

0.9	25	12	334	33	138	12	472	45	0.912959381
0.9	45	2	359	8	45	105	404	113	0.781431335
0.9	45	4	267	100	144	6	411	106	0.794970986
0.9	45	8	325	42	143	7	468	49	0.905222437
0.9	45	12	335	32	138	12	473	44	0.914893617

Naïve Bayes Appendix:

Training %	Ham correct	Ham incorrect	Spam Correct	Spam Incorrect	Total Correct	Total Incorrect	Accuracy
0.1	3265	39	1153	197	4418	236	0.949291
0.25	2728	26	1039	86	3767	112	0.971127
0.5	1810	26	707	43	2517	69	0.973318
0.75	906	12	354	21	1260	33	0.974478
0.9	365	2	139	11	504	13	0.974855

KNN Appendix:

	Accuracy on Test Data
.25 Training %, 1 NN	0.86825
.25 Training %, 5 NN	0.84119
.25 Training %, 10 NN	0.81747
.25 Training %, 25 NN	0.8056
.50 Training %, 1 NN	0.8894
.50 Training %, 5 NN	0.8576
.50 Training %, 10 NN	0.8383
.50 Training %, 25 NN	0.805104
.75 Training %, 1 NN	0.914
.75 Training %, 5 NN	0.876

4 Email Spam Classification Case Study 40 / 40

✓ **- 0 pts** **No major errors**

- 5 pts partially missing classifier or mistake

- 10 pts missing an entire classifier or wrong implementation

- 20 pts Code not present or major part missing in code

- 25 pts Incorrect or not present 4 ii)

- 5 pts Incorrect/invalid/missing justification for 4 iii)

- 10 pts Missing performance graphs for 4 iii)

- 2 pts Missing justified final decision on best model for 4 iii)

- 2.5 pts Partially incorrect justification for 4 iii)

- 5 pts Partially missing performance analysis or graphs

- 18 pts Code Copied From External Sources Adjustment

next time when submitting file please make sure you are not submitting the IDE's config files

5 Adjustments **0 / 0**

- ✓ **- 0 pts** **No adjustments**
- **- 3 pts** Pages assigned incorrectly
- **- 3 pts** Code submitted incorreclty
- **- 115 pts** HOLD (please contact the instructor directly to release the hold)