Name : Kapil Kumar Chhipa
Roll.No : M180265CA
Sub : Machine Learning

(4)

	A	B	C	
A	80	9	11	100
B	73	11	16	100
C	0	42	18	60
				260

$$\text{Accuracy} = \frac{\text{total no. of correct Prediction}}{\text{total dataset}}$$

$$\frac{80+11+18}{260} \qquad \frac{109}{260}$$

$$= 0.4192 \quad \Delta$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

P(A) = Precision(A)

$$P(A) = \frac{80}{153} = 0.5228$$

$$\text{Precision}(B) = \frac{11}{62} = 0.1774$$

$$\text{Precision}(C) = \frac{18}{45} = 0.4$$

Average Precision

$$\frac{P(A)+P(B)+P(C)}{3} = \frac{1.1002}{3} = 0.3667$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$R(A) = \frac{80}{100} = 0.8$$

$$R(B) \quad \frac{11}{100} = 0.11$$

$$R(C) = \frac{18}{60} = 0.3$$

Average Recall =

$$\frac{R(A)+R(B)+R(C)}{3}$$

$$\frac{0.8+0.11+0.3}{3} = \frac{1.21}{3} = 0.40$$

$$F1\ score = \frac{2 \times 0.3667 \times 0.4033}{0.3667 + 0.4033}$$

$$\frac{0.2957}{0.77} = 0.3840$$

$$F1\ score = 0.3840$$

(1)

Computer	P_1	P_2	Decision
C1	6	7	Yes
C2	8	5	Yes
C3	4	6	No
C4	10	9	No
C5	9	7	Yes

$P_1 = 5$, $P_2 = 2$ $k = 4$

distance equation is

$$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2} - - -$$

for C1 $\sqrt{(6-5)^2+(7-2)^2}$

$\sqrt{1+25} = \sqrt{26} = 5.0990$

for C_2 $\sqrt{(8-5)^2+(5-2)^2}$

$\sqrt{9+9} = \sqrt{18} = 4.2426$

for C_3 $\sqrt{(4-5)^2+(6-2)^2}$

$\sqrt{1+16} = \sqrt{17} = 4.12310$

for C5 $\sqrt{(9-5)^2+(7-2)^2}$

$\sqrt{16+25} = \sqrt{41} = 6.4031$

~~for C5 $\sqrt{(9-5)^2+(7-$~~

for C4 $\sqrt{(10-5)^2+(9-2)^2}$

$\sqrt{25+49} = \sqrt{74} = 8.6023$

C1	5.0990
C2	4.2426
C3	4.12310
C4	8.6023
C5	6.4031

find 4 ~~Closest~~ smallest results

~~C2~~	C1	5.0990	Yes
~~C2~~	C2	4.2426	Yes
	C3	4.12310	No
	C5	6.4031	Yes

The decision will be Yes by ~~KK~~ KNN algorithm. because we have 3 Yes and 1 No.

PAGE NO.:
DATE: / /

It has 3 neighbor taking decision Yes and 1 neighbor taking decision as NO.

(2) (a)

Since the hypothesis performs well (has low error) on the training set, it is suffering from high variance.

→ Try using a smaller set of features;

The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Reducing the feature set will ameliorate the overfitting and help with the variance problem.

→ Get more training example: The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Adding more training data will increase the complexity of the training set and help with the variance problem.

→ Try increasing the regularization parameter
A. The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Increasing the regularization parameter will reduce overfitting and help with the variance problem

1 (2)

(b) Since the hypothesis performs poorly (after reversing the (a) scenario) on the training set, It is suffering from high bias.

→ Try decreasing the regularization parameter λ: Decreasing the regularization parameter will improve the high bias problem and may improve the performance on the training set.

→ Try adding polynomial features:
The poor performance on both the training and test sets suggests a high bias problem. Adding more complex features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.

→ Try to obtain and use additional features:
The poor performance on both the training and test sets suggests a high bias problem. Using additional features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.

(3)

(i) ["The Indian Rebellion of 1857 began as a mutiny of sepoys of the East India Company's army on 10 May 1857, in the town of Meerut, and soon escalated into other mutinies and civilian rebellions largely in the upper Gangetic plain and central India, with the major hostilities confined to present-day Uttar Pradesh, Bihar, northern Madhya Pradesh, and the Delhi region."]

Note: I am keeping one line for clearity.

(11)

['The', 'Indian', 'Rebellion', 'of', '1857', 'began', 'as', 'a', 'mutiny', 'of', 'sepoys', 'of', 'the', 'East', 'India', 'Company', "'s", 'army', 'on', '10', 'May', '1857', ',', 'in', 'the', 'town', 'of', 'Meerut', ',', 'and', 'soon', 'escalated', 'into', 'other', 'mutinies', 'and', 'civilian', 'rebellions', 'largely', 'in', 'the', 'Upper', 'Gangetic', 'plain', 'and', 'central', 'India', ',', 'with', 'the', 'major', 'hostilities', 'confined', 'to', 'present-day', 'Uttar', 'Pradesh', ',', 'Bihar', ',', 'northern', 'Madhya', 'Pradesh', ',', 'and', 'the', 'Delhi', 'region', '.']

Stopwords Removal

(iii) Note: i am assuming words are case-sensitive.

['The', 'Indian', 'Rebelian', '1857', 'began', 'muting', 'sepoys', 'East', 'India', 'Company', "'s", 'army', 'to', 'May', '1857', ',', 'town', 'Meerut', ',', 'Soon', 'escalated', 'mutinies', 'civilian', 'rebelions', 'largely', 'upper', 'Gangetic', 'Plain', 'central', 'India', ',', 'major', 'hostilities', 'combined', 'Present-day', 'Uttar', 'Pradesh', ',', 'Bihar', ',', 'northern', 'Madhya', 'Pradesh', ',', 'Delhi', 'region', '.']

(IV)

[('The', 'DT'), ('Indian', 'JJ'), ('Rebellion', 'NNP'), ('of', 'IN'), ('1857', 'CD'), ('began', 'VBD'), ('as', 'IN'), ('a', 'DT'), ('mutiny', 'NN'), ('of', 'IN'), ('sepoys', 'NNS'), ('of', 'IN'), ('the', 'DT'), ('East', 'NNP'), ('India', 'NNP'), ('Company', 'NNP'), ("'s", 'POS'), ('army', 'NN'), ('on', 'IN'), ('10', 'CD'), ('May', 'NNP'), ('1857', 'CD'), (',', ','), ('in', 'IN'), ('the', 'DT'), ('town', 'NN'), ('of', 'IN'), ('Meerut', 'NNP'), (',', ','), ('and', 'CC'), ('soon', 'RB'), ('escalated', 'VBD'), ('into', 'IN'), ('other', 'JJ'), ('mutinies', 'NNS'), ('and', 'CC'), ('civilian', 'JJ'), ('rebellions', 'NNS'), ('largely', 'RB'), ('in', 'IN'), ('the', 'DT'), ('Upper', 'JJ'), ('Gangetic', 'NNP'), ('Plain', 'NN'),

('and', 'cc'), ('central', 'JJ'), ('India', 'NNP'),
(',', ','), ('with', 'IN'), ('the', 'DT'),
('major', 'JJ'), ('municipalities', 'NNS'),
('confined', 'VBN'), ('to', 'TO'), ('Present-
day', 'JJ'), ('Uttar', 'NNP'), ('Pradesh',
'NNP'), (',', ','), ('Bihar', 'NNP'),
(',', ','), ('northern', 'JJ'), ('Madhya',
'NNP'), ('Pradesh', 'NNP'), (',', ','), ('and',
'CC'), ('the', 'DT'), ('Delhi', 'NNP'),
('region', 'NN'), ('.', '.')]

(v) Tree ('S', [('The', 'DT'), ('Indian', 'JJ'), ('Rebellion', 'NNP'), ('of', 'IN'), ('1857', 'CD'), ('began', 'VBD'), ('as', 'IN'), Tree('NP', [('a', 'DT'), ('mutiny', 'NN')]), ('of', 'IN'), ('sepoys', 'NNS'), ('of', 'IN'), ('the', 'DT'), ('East', 'NNP'), ('India', 'NNP'), ('Company', 'NNP'), ("'s", 'POS'), Tree('NP', [('army', 'NN')])),

grammer is

"NP: {<DT>? <JJ>* <NN>}"