Assignment 5

Bella Fu jf70

Task 1

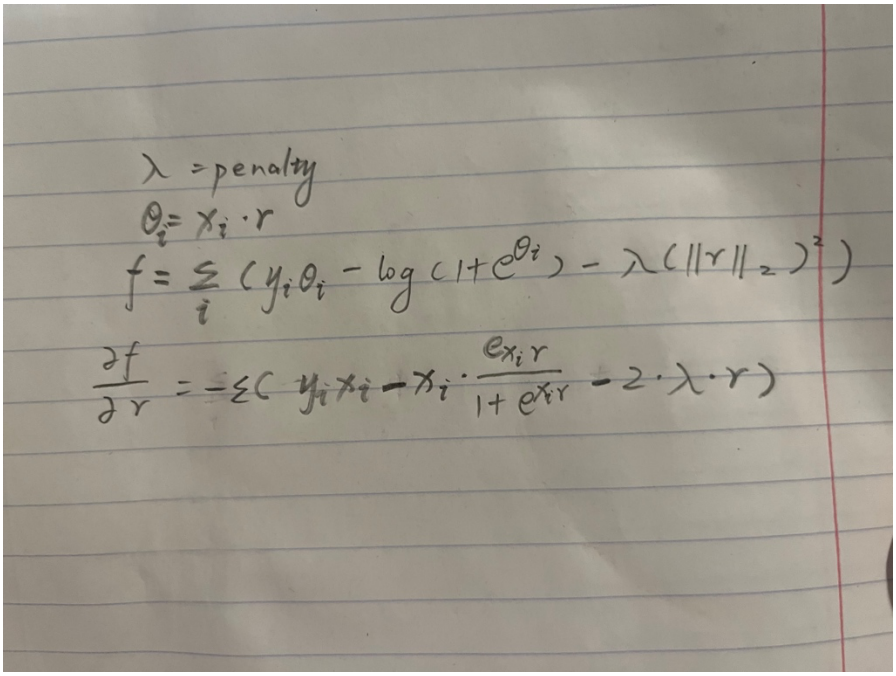
Data option: Training Data Set

Result:
448
2
514
3167
652

Task 2

Data option: Small Data Set

(a) Gradient update formula



(b) Fifty words with largest regression coefficients

hca
fca
satisfied
pursuant
applicant
respondent
relevantly
clr
pty
alr
tribunal
submissions
fcafc
relevant
relation
circumstances
affidavit
interlocutory
whether
reasons
proceeding
jurisdictional
honour
appellant
jj
multicultural
respondents
gummow
mr
error
consideration
fmca
respect
ltd
relied
gaudron
subsection
notice
proceedings
delegate

affidavits
leave
visa
costs
hearing
contravention
sought
conduct
fcr
solicitors

Task 3

Data option: Small Data Set

(1) *F1 score result*: 0.87318635016
(2) *Cut off*: 0.2
(3) *Potential reason of why my model was fooled:*

Looking at these three false-positive context, they have many contents about the legal systems, therefore they have many words in common with our target category. For example, 'jurisdictional', 'proceeding'.
That might be one of the reasons why my classifier recognized them as 'positive'.