COMP 543, Tools and Models for Data Science

Chengyin Liu, cl93

Assignment #5

I tried both the small data and the big data, enjoying training.

################

#Task 1

1. Train on small data 'TestingDataOneLinePerDoc.txt'

The frequency position of the words "applicant", "and", "attack", "protein", and "car":

603

2

514

3683

634

####################

2. Train on big data 'TrainingDataOneLinePerDoc.txt'

The frequency position of the words "applicant", "and", "attack", "protein", and "car":

447

2

513

3166

651

###################

###################

#Task 2

The logistic regression lost function with regularization is

$$J(\theta) = -\frac{1}{m}\{\sum_{i=1}^{m}[y^{(i)}log(h\theta(x^{(i)})) + (1-y^{(i)})log(1-h\theta(x^{(i)}))]\} + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

,

Where α is the learning rate and λ is the regularization parameter.

And the gradient update formula with regularization I use is as bellow,

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \qquad \text{for } j = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right) + \frac{\lambda}{m}\theta_j \qquad \text{for } j \geq 1$$

,

Then we repeat the weights as

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

.

###############

1. Train on small data ‘TestingDataOneLinePerDoc.txt’

The fifty words with the largest regression coefficients:

applicant

clr

pty

fcr

hca

alr

respondent

relevantly

fca

tribunal

submissions

fcafc

gummow

affidavit

pursuant
proceeding
application
satisfied
appellant
respondents
submits
gaudron
costs
amp
proceedings
reasons
jurisdictional
nswlr
relevant
affidavits
magistrate
interlocutory
mr
hearing
notice
arguable
multicultural
contravention
relation
evidence
circumstances
honour
acsr
magistrates

applicants

paragraphs

contraventions

orders

erred

ltd

###############

2. Train on big data ‘TrainingDataOneLinePerDoc.txt’

The fifty words with the largest regression coefficients:

applicant

hca

clr

fcr

respondent

fca

pty

relevantly

tribunal

fcafc

pursuant

appellant

proceeding

affidavit

alr

submissions

gummow

application

satisfied

notice

respondents

interlocutory

submits

nswlr

costs

proceedings

jurisdictional

magistrate

mr

erred

contravention

applicants

reasons

affidavits

arguable

multicultural

circumstances

relevant

appeal

relation

discretion

amp

claim

acsr

paragraphs

contentions

hearing

fmca

orders

solicitors

################

################

#Task3

1, Test on small data ‘SmallTrainingDataOneLinePerDoc.txt’

F1 score: 0.961038961038961

I looked at three false positives: '3617964', '36456184', and '457608', they are all talking about the legal systems. As for example the document '3617964' showed bellow, many words with the top 50 coefficients, like ‘applicant’, ‘evidence’, ‘proceeding’, and so on, appear in these documents. I think this is why my model thinks they also belong to ‘AU’ documents.

[u' Bankruptcy and Insolvency Act The Bankruptcy and Insolvency Act ("BIA") is the statute that regulates the law on bankruptcy and insolvency in Canada. It governs bankruptcies, consumer and commercial proposals, and receiverships in Canada. It also governs the Office of the Superintendent of Bankruptcy, a federal agency responsible for ensuring that bankruptcies are administered in a fair and orderly manner. Purpose and scope. The nature of the Act within Canada\'s legal framework governing insolvency was described by the Supreme Court of Canada in "Century Services Inc. v. Canada (Attorney General)": With certain exceptions, the "BIA" covers a wide range of entities: The Act governs bankruptcy proceedings, which are invoked: The Act also governs receivership proceedings. Receivers may be appointed by a secured creditor under the terms of a general security agreement (where the debtor voluntarily agrees), or by the court where a secured creditor: Provision is also made for dealing with cross-border insolvencies and the recognition of foreign proceedings. Relationship with provincial law. Several notable cases known as the "bankruptcy quartet" stand for the following propositions about how the Act interacts with provincial legislation: In December 2013, the Ontario Court of Appeal held that, as a result of the above, a discharge from bankruptcy served as a release from all claims in bankruptcy, including tolls charged by 407 ETR, and Ontario\' s "Highway 407 Act, 1998" could not bar a discharged bankrupt from renewing his license plates upon payment of normal annual fees. Leave to appeal the decision was granted by the Supreme Court of Canada in May 2014. History an d development. Consolidation of pre-Confederation legislation. No specific legislation on bankruptcy and insolvency previously existed in New Brunswick and Nova Scotia. Bankruptcy process. Protective provisions. A secured creditor cannot enforce security on the business assets of an insolvent person without having given 10 days\' advance notice in the prescribed form and manner. No person may terminate or amend \u2014 or claim an accelerated payment or forfeiture of the term under \u2014 any agreement, including a security agreement, with a bankrupt individua l by reason only of the individual\u2019s bankruptcy or insolvency. Similar provision is made with respect to any insolvent person upon filing a notice of intention or a proposal. A notice of intention, a Division I proposal, or a Division II proposal, will automatically create a stay of proceedings and "no creditor has any remedy against the debtor or the debtor\u2019s property, or shall commence or continue any action, execution or other proceedings, for the recovery of a claim provable in bankruptcy". Similar provision is also made on the

bankruptcy of any debtor. Directors of insolvent companies that have filed a notice of intention or a proposal have similar protection. Suspension of attachments. S. 70(1) of the "BIA" provides that bankruptcy orders and assignments take precedence over "all judicial or other attachments, garnishments, certificates having the effect of judgments, judgments, certificates of judgment, legal hypothecs of judgment creditors, executions or other process against the property of a bankrupt," but that does not extend to: The Ontario Court of Appeal has ruled that, in the case of a "requirement to pay" under the "Income Tax Act" (Canada) that was issued after a notice of application to appoint a receiver (but before the court heard the application), supported by an "ex parte" "jeopardy order" issued by the Federal Court of Canada under s. 225.1(1) of that Act, the "requirement to pay" was considered to have been completely executed on the date of its issue, and thus took precedence over other creditors\' claims. Settlement of the insolvent person\'s estate. The trustee/receiver must first realize the amount of the proceeds from the proper ty that is available for payment to the different classes of creditors, and different rules apply according to the type of proceeding. They are summarized as follows: The estate is then settled, using the priority of claims outlined in the "BIA". Creditors. The resulting amount available from the estate is distributed to the creditors in the following order of priority (with each class/subclass paid in full before proceeding to the next): There are several important notes to consider in assessing the above priorities: Every creditor must prove his claim and a creditor who does not prove his claim is not entitled to any distribution of the proceeds from bankrupt\u2019s estate. The claim must be delivered to the trustee in bankruptcy and the trustee in bankruptcy must examine every proof of claim and can request further proof. The trustee may disallow, in whole or in part, any claim of right t o a priority under the BIA or security. Generally, the test of proving the claim before the trustee in bankruptcy is very low, and a claim is proved unless it is too "remote and speculative". The rationale for such a low test is to discharge as many claims as possible to allow the bankrupt to make a fresh start after the discharge. Creditors also have the ability, with the approval of the court, to take over a cause of action that the trustee has decided not to pursue. Some liabilities are not released upon discharge. Directors and parties related to the bankrupt may still be held personally liable for certain tax debts, and directors can be held accountable for other liabilities. Preferences and transfers at undervalue. In 2009, the BIA was amended to reform the rules relating to setting aside any preferences, or transfers at undervalue, occurring before the initial bankruptcy event: Recovery actions under ss. 95 and 96, as for other recovery actions with respect to collections, can only be initiated by the trustee, even when they may be of benefit only to a secured creditor (unless creditors seek court approval under s. 38 to pursue the matter directly). The BIA already empowered the court to inquire into circumstances where a bankrupt corporation had paid cash dividends or redeemed shares where the corporation was insolvent, or where the transactions made it so, during the 12 months prior to its bankruptcy. In that regard, S. 95(2) provides that, where a preference is given, the fact that it may have been given under pressure is irrelevant. However, the courts have ruled that a payment may withstand challenge by a trustee where it is made in furtherance of a reasonable business imperative. Key actors in the procedure. Bankruptcy court. The provincial Superior Courts have "such jurisdiction at law and in equity" as will enable them to exercise bankruptcy process under the BIA. The decisions of the court are enforceable in the courts of other Canadian provinces and all courts and the officers of al

l courts must act and co-operate in all bankruptcy matters. Appeal from the court\'s orders lies to the provincial Court of Appeal where: Registrars of the provincial Superior Courts have significant powers in relation to procedural matters, unopposed proceedings and in other matters under the Act. Office of the Superintendent of Bankruptcy. The Office of the Superintendent of Bankruptcy ("OSB") is designed to supervise the administration of all estates and matters to which the BIA applies. It grants licenses for the trustees in bankruptcy, inspects investigates bankruptcy estates, reviews the conduct of the trustees in bankruptcy and the receivers, and examines trustee \u2019s accounts, receipts, disbursements and final statements. It has specific powers to intervene in any matter or proceeding in court as if the OSB were a party thereto, as well as to issue directives providing official interpretation of the bankruptcy process to the trustees in bankruptcy and the receivers. Trustee in bankruptcy. Trustees either individuals or corporations are licensed by the Superintendent, and are appointed to administer an e state by virtue of the assignment, bankruptcy order or proposal that has been filed. By special resolution, the creditors of the estate may appoint or substitute another licensed trustee to assume the role. A trustee is not bound to accept an appointment, but, once appointed, he must perform all duties that are legally required until his discharge or removal. Otherwise, any licensed trustee can be appointed to act, subject to the following constraints: The trustee acts as receiver for all the estate\'s property, and is entitled to see its books and records. All moneys he receives must be deposited into a separate trust account. When required, he is obliged to report on the estate\'s condition, moneys on hand, and property remaining unsold. He is not obliged to continue the business of the bankrupt, where there is no good business case for doing so. When he has completed the duties required of him for administering the estate, he shall apply to the court for a discharge, but any interested person may file an objection to having the discharge take place. All property of the bankruptcy vests in the trustee from the date of the bankruptcy, and the trustee may register a bankruptcy order against any real property in which the bankrupt has any interest or estate. The courts have held that trustees should clearly communicate to the bankrupt their intent to make a claim against the non-exempt equity in the bankrupt\'s property at the time of the assignment into bankruptcy. Failure to do so may result in: The Superintendent may undertake conservatory measures in or der to protect an estate, as well as the rights of the creditors and debtors, in specified circumstances: Inspectors. At the first meeting of the creditors, up to five individuals may be appointed to be inspectors of the estate (except where the creditors decide that that is not necessary). No inspector may be appointed if he is a party to any contested action or proceeding against the estate. Where the value of an individual debtor\u2019s property is under $15,000, inspectors are not appointed (except where the creditors decide otherwise). The trustee is required to obtain the inspectors\' permission before carrying out many of his responsibilities, such as the sale of property of the estate, the institution or defending of actions relating to the property of the bankrupt, settling any debts owing to the bankrupt and exercising trustee\u2019s discretion in retaining and assigning bankrupt\u 2019s contracts. The inspectors must give their approval to the final statement of receipts and disbursements and trustee\u2019s fees. Inspectors have a fiduciary duty to the creditors and should be impartial though acting in their interest. They should supervise the trustee\u2019s compliance with the BIA and the Superintendent\'s directives and may apply for the removal of the trustee. Receivers. The

receiver must do what "practicality demands" to preserve the assets and must not go beyond what is necessary in the circumstances. Interim receivers. The court may appoint an interim receiver: In the first case, the applicant must give an undertaking with respect to the debtor\'s legal rights, and to damages in the event of the application being dismissed. The interim receiver can take conservatory measures and dispose of perishable property in order to comply with the order of the court, but the receiver cannot otherwise unduly interfere with the bankrupt in the carrying on of the debtor\u2019s business. In the latter two cases, the court can only make the appointment if it is shown that it is necessary for the protection of the debtor\'s estate, or in the interest of the creditor(s). The courts have set out the following factors to be considered in exercising discretion on whether to appoint an interim receiver: ']

###############

2. Test on big data ‘TestingDataOneLinePerDoc.txt’

F1 score: 0.9826897470039947

I looked at three false positives: '41034992', '37359041', '41766278'

They are also talking about the legal systems. Many words with the top 50 coefficients appear in these documents just as in the small test dataset. So, my model thinks they also belong to ‘AU’ documents.

For example, the ‘41034992’ document showed bellow, it about Security, Court and government laws, so it is very close in content to our positive training samples.

[u' Khala v Minister of Safety and Security Khala v Minister of Safety and Security is an important case in South African law. Facts. The plaintiff had instituted action against the defendant for damages arising from an alleged unlawful arrest and detention. The defendant considered that the police docket relating to such arrest was privileged, and listed it accordingly in his discovery affidavit. The plaintiff then launched the application "in casu" for an order directing the defendant to make the police docket available for inspection and copying. Judgment. "Amicus curiae". After the postponement of the matter, from May 19, 1994, to June 13, 1994, the court (with the consent of the parties, and having obtained the approval of the Judge President) appointed Prof E. Mureinik of the Law Faculty of the University of the Witwatersrand to act as "amicus curiae". Prof Mureinik filed heads of argument, provided copies of authorities, and made oral submissions after those of the plaintiff and defendant had been concluded. Counsel for the plaintiff and defendant thereafter replied to his submissions. He was not remunerated for this work. His assistance was "invaluable to the Court and was greatly appreciated." Merits. Having set out the provisions of the Constitution relevant to the issue, the court considered the principles applicable in interpreting provisions of a constitution. Fundamental-rights provisions call for a generous and purposive interpretation. The meaning of a right is to be gathered from a consideration of the interests it was intended to protect. The principles of interpretation contained in section 35 of the Constitution are

also applicable. The court then turned to an examination of Plaintiff\u2019s rights under section 23. It was accepted that the information contained in the police docket was held by an organ of state, and was information relevant to the protection of the plaintiff\u2019s rights to freedom and security. The defendant, however, placed in issue that the information was \u201crequired\u201d by the plaintiff. After a survey of judicial interpretations of the word \u201crequired,\u201d the Court concluded that, in its context in section 23, whether information is \u201crequired\u201d in any particular case is a factual question. This raised the issue of whether it was intended that section 23 should serve as a discovery measure in litigation between the government and another. The defendant contended that it was not, and that section 23 should be viewed as analogous to the freedom-of-information statutes enacted in various other countries. The court held that such an analogy is not proper. Section 23 does not give the public a general right of access to information. It confers on individuals a right of access to information which is required for the exercise or protection of a right. Then and only then does the State have an obligation to provide access. To resist a claim for information, the State would have to satisfy the requirements of section 33 in each particular case. The Court concluded that it was appropriate to use section 23 to obtain discovery of documents, and that the plaintiff was entitled in terms of section 23 to the information in the police docket. The court then turned to a consideration of the defendant\u2019s reliance on section 33(1). This resolved into a question of whether docket privilege is reasonable and justifiable in an open and democratic society based on freedom and equality. (In regard to the requirement of section 33(1)(b)\u2014that a limit \u201cshall not negate the essential content of a right\u201d\u2014the court found that docket privilege did not negate the essential content of the section 23 right.) The court held that Defendant bore the onus to establish this according to the civil standard of proof. Before turning to a consideration of the affidavit of the Commissioner of Police in support of the defendant\u2019s contentions, the Court set out the nature of docket privilege. Prior to 1954, only three categories of privilege were recognised: "R v Steyn" conferred a privilege on police dockets. In creating this general privilege, appeal was made to considerations of public policy. Thereafter, judicial precedent extended the privilege and added to it the principle, taken from legal professional privilege, of \u201conce privileged, always privileged.\u201d In his affidavit, the Commissioner of Police advanced reasons justifying the withholding of information contained in the police docket in a criminal case. These included In assessing the cogency of such a justification, the court considered that a starting point must be an acceptance that the administration of justice and the maintenance of social peace and order is a fundamental public interest. This dictated that, as a matter of public policy, some information (such as the identity of informers) should be subject to privilege and not be disclosed. The defendant\u2019s case, however, was that, because some of the information might be privileged, all the information should be withheld. This approach did not deal with whether the non-disclosure of unprivileged information in the docket was justifiable in an open and democratic society based on freedom and equality. To determine this question, the parties had been requested to inform the court of the practice in other countries. By agreement the parties, jointly requested Professor T Geldenhuys of the University of South Africa to prepare a memorandum setting out research conducted by him on this question. Such research covered discovery practice in criminal cases in Canada, New Zealand, the United States of America, England and Australia. The findings

are set out in the court\'s judgment. The Court concluded from that survey that, in general, the trend in those societies was towards an expansion of the accused\u2019s rights of access to information in the possession of the prosecution. In the court\u2019s view, South African law should fall in line with that international trend. Policy considerations in favour of disclosure of unprivileged information outweigh those against disclosure. Any residual doubt on where the balance lay was removed by taking into account the fundamental right of equality before the law (section 8(1)), the right to a fair trial (section 25(3)) and the presumption of innocence (section 25(3)(c)). The Court concluded that Defendant had not established that the privilege hitherto attaching to the police docket in respect of its unprivileged contents was reasonable and justifiable in an open and democratic society based on freedom and equality. The Court rejected the claim of docket privilege and granted the defendant leave to file a supplementary discovery affidavit disclosing, in the first schedule, the information in respect of which no privilege was claimed and in the second schedule, that information in respect of which privilege was claimed. ']