

Exemple 2 :

Objectif :

Déterminer une probabilité d'avoir les cheveux longs ou courts

Principe :

Les caractéristiques renseignées pour chaque individu de la base de donnée sont:

- Le sexe
- L'année de naissance
- Le nombre d'heure consacrées à se coiffer par semaine
- La taille des cheveux (0 pour court, 1 pour long)

On entraîne le réseau.

Pour un individu dont on connaît le sexe l'année de naissance et le nombre d'heures consacrées à se coiffer on détermine la probabilité qu'il ait les cheveux courts ou longs.

Caractéristique du réseau de neurone :

Type:

Régression linéaire

Le réseau va agir comme une régression linéaire mais sur plusieurs dimensions (âge, sexe, etc ..)

Dimension :

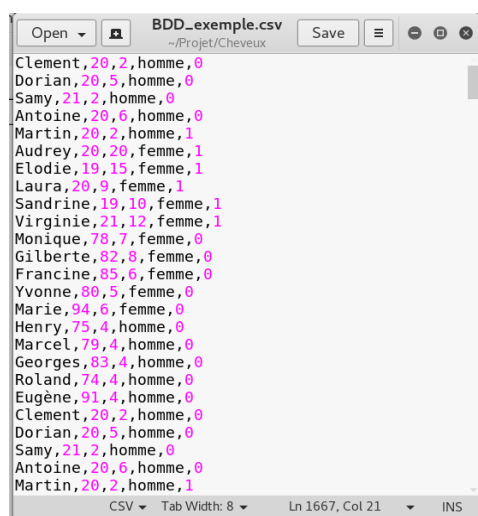
3 couches de taille 128 (à partir de 4 couches le réseau devient inefficace car les informations traitées sont « simples » et multiplier les chemins possibles diminue la qualité du réseau dans notre cas)

Fonction d'activation :

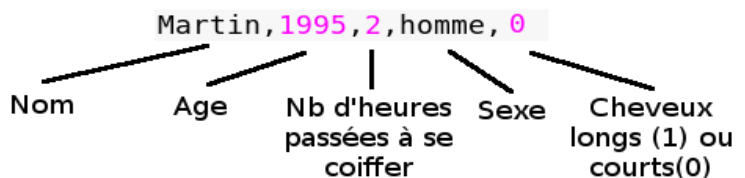
Softmax : $r_i(a) = e^{a_i} / \sum_j e^{a_j}$ (équivalent à la sigmoïde en 2 dimensions)

Naturellement associée au principe de probabilité.

Entrées :



| Nom | Age | Nb d'heures passées à se coiffer | Sexe | Cheveux longs (1) ou courts(0) |
|----------|-----|----------------------------------|-------|--------------------------------|
| Clement | 20 | 2 | homme | 0 |
| Dorian | 20 | 5 | homme | 0 |
| Samy | 21 | 2 | homme | 0 |
| Antoine | 20 | 6 | homme | 0 |
| Martin | 20 | 2 | homme | 1 |
| Audrey | 20 | 20 | femme | 1 |
| Elodie | 19 | 15 | femme | 1 |
| Laura | 20 | 9 | femme | 1 |
| Sandrine | 19 | 10 | femme | 1 |
| Virginie | 21 | 12 | femme | 1 |
| Monique | 78 | 7 | femme | 0 |
| Gilberte | 82 | 8 | femme | 0 |
| Francine | 85 | 6 | femme | 0 |
| Yvonne | 80 | 5 | femme | 0 |
| Marie | 94 | 6 | femme | 0 |
| Henry | 75 | 4 | homme | 0 |
| Marcel | 79 | 4 | homme | 0 |
| Georges | 83 | 4 | homme | 0 |
| Roland | 74 | 4 | homme | 0 |
| Eugène | 91 | 4 | homme | 0 |
| Clement | 20 | 2 | homme | 0 |
| Dorian | 20 | 5 | homme | 0 |
| Samy | 21 | 2 | homme | 0 |
| Antoine | 20 | 6 | homme | 0 |
| Martin | 20 | 2 | homme | 1 |



Programme : Voir programme 2

Résultats :

(Coby, 1992 , 7 , femme) → 83%

(Coby , 1992 , 15 , femme) → 99.6%

(Coby , 1996 , 4 , homme) → 35%

(Coby , 1996 , 4 , femme) → 52%

(Coby, 1947, 7 , femme) → 83%

Points de satisfaction :

- Le caractère sexe est déterminant.
- Une augmentation du nombre d'heure consacrées à se coiffer augmente la probabilité d'avoir les cheveux longs.

Points à améliorer :

- La caractéristique année de naissance n'est pas déterminante

Si on transforme toutes les années de naissance en âge (exemple : 1947 devient 80) alors le nombre d'heure consacrée à se coiffer est éclipsé par l'âge du cobaye, qui devient la deuxième caractéristique la plus déterminante du réseau derrière le sexe.

Conclusion pour cet exemple:

La probabilité donnée pour chacun des cobayes est cohérente et le réseau fonctionne correctement.

Nous avons constaté que :

- Le nombre de couche doit être adapté à la complexité des informations transmises
- La distribution des données est importante dans la détermination des coefficients du réseau

Exemple 3 :

Objectif :

Générer un poème.

Principe :

On crée un réseau neurone adapté.

On nourrit le réseau de neurone avec un recueil de poème de très grande taille.

On génère une séquence de lettres dont l'enchaînement est généré par le réseau de neurone.

Caractéristiques du réseau de neurone :

Type:

Notre réseau sera du type LSTM (Long Short-Term Memory).

Il appartient à la famille des réseaux récurrents aussi appelé RNN (Recurrent Neural Network) qui permettent de traiter des informations temporelles.

Exemples de réseau utilisant des informations temporelles :

- La prévision météo
- La reconnaissance vocale

Exemple non temporelle:

- La classification d'individu

Dans notre exemple, les informations à traiter sont des poèmes. L'enchaînement des lettres respecte certaines règles et la génération d'une nouvelle lettre ne peut se faire sans tenir compte des précédentes.

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Dimension :

3 couches de taille 512

Fonction d'activation :

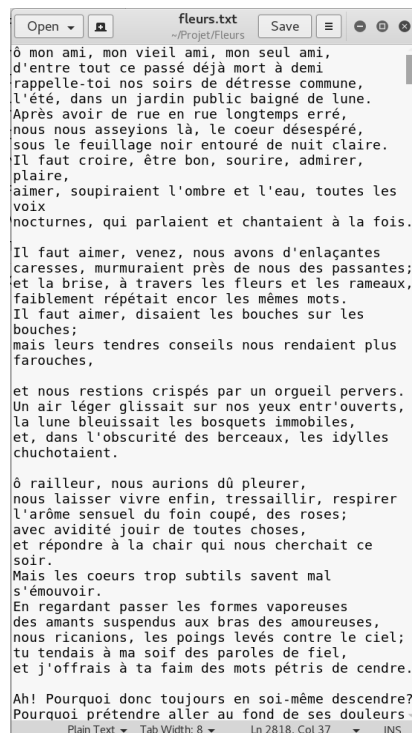
Softmax : Il est une nouvelle fois question de probabilité, ici nous travaillons avec une probabilité d'apparition d'une lettre en fonction des précédentes.

Graine :

Le réseau de neurone construit une matrice de probabilité en s'appuyant sur les lettres précédentes. Le programme ne sait pas « choisir » les toutes premières lettres qui permettront de démarrer la génération. Ces quelques premières lettres constituent la graine.

Ici, on choisit aléatoirement un court passage du recueil qui servira d'amorce au programme.

Entrées :



Open Save
fleurs.txt
~Projet/Fleurs

ô mon ami, mon vieil ami, mon seul ami,
d'entre tout ce passé déjà mort à demi
rappelle-toi nos soirs de détresse commune,
l'été, dans un jardin public baigné de lune.
Après avoir de rue en rue longtemps erré,
nous nous asseyions là, le cœur désespéré,
sous le feuillage noir entouré de nuit claire.
Il faut croire, être bon, sourire, admirer,
plaire,
aimer, soupiraient l'ombre et l'eau, toutes les
voix
nocturnes, qui parlaient et chantaient à la fois.

Il faut aimer, venez, nous avons d'enlçantes
caresses, murmuraient près de nous des passantes;
et la brise, à travers les fleurs et les rameaux,
faiblement répétait encor les mêmes mots.
Il faut aimer, disaient les bouches sur les
bouches;
mais leurs tendres conseils nous rendaient plus
farouches,

et nous restions crispés par un orgueil pervers.
Un air léger glissait sur nos yeux entr'ouverts,
la lune bleuissait les bosquets immobiles,
et, dans l'obscurité des berceaux, les idylles
chuchotaient.

ô railleur, nous aurions dû pleurer,
nous laisser vivre enfin, tressaillir, respirer
l'arôme sensuel du foin coupé, des roses;
avec avidité jouir de toutes choses,
et répondre à la chair qui nous cherchait ce
soir.
Mais les cœurs trop subtils savent mal
s'émouvoir.
En regardant passer les formes vaporeuses
des amants suspendus aux bras des amoureuses,
nous ricanions, les poings levés contre le ciel;
tu tendais à ma soif des paroles de fiel,
et j'offrais à ta faim des mots pétris de cendre.

Ah! Pourquoi donc toujours en soi-même descendre?
Pourquoi prétendre aller au fond de ses douleurs-

Plain Text Tab Width: 8 Ln 2818, Col 37 INS

Programme : Voir programme 3

Résultats :

La température permet de jouer sur l'excentricité de la séquence générée. Plus elle est faible et plus la génération est déterministe.

Après 15 minutes d'entraînement :

```
TEST:
Température =1 :
mbeau d'Orphée.
La terreovt mlaebeusug it se s i eaeaeae pe . seoalt rc Dpnsle
ma'isn: dt lRpb edué r epnP eoeitrEtuirevt : s eev,aee;uair stio tlt esstle
i N u lahuotscldhlt e ceeufP nen v
o olvveefateioeoruessmte dsrahesrell .rnli erss gse tiezn !n'iuepnmeeiri'oelo, c modnuon ct nie i;'drhba
srcl r olLa nerc

Température =0.5 :
mbeau d'Orphée.
La terrer seda o s en enl e ara ee sn e ss um a s ee s o er mul e s u i es
rsos ser teoe u l i
oe t a s e s s r esme o sr des on s r oerash nos ee f o etoo ae e
dsr ea s te d a a r sene sarl e rs t ast eeeoe eeu'ece u esoeouos etosee in d eeasoeel
```

Après 3h d'entraînement :

TEST:

Température =1 :

re ce soir dans le logis darcleant les coufanchon de ciedes-en unt puseltatt et sen
eux cau pomi centte d'éraorse qve marge marse, plubripn
oifapront coitsie er be l'étente
dusucasti tentie éconcit moicrite et plais l'eurair serte;
-utit oupeux das saurrérert,
et dauns ce qvacn enge,
l'atpis de desn

Température =0.5 :

re ce soir dans le logis le la tain et pont le perde en faisurene eu des le chaneir de conte des coel de lans
et les en te soir de sartes des corges et de moerte
et da toin
et l'angre et tor en feur murer de sont en de chaute et la vens vass féchant dus l'aisein dant se mecbe ront
des dans ene le pures de léle des le noure

Points de satisfaction :

- La structure de mot et de phrases est reproduite dans la séquence générée.
- La séquence respecte un enchaînement voyelle consonnes lisibles à l'œil et à consonance française
- On retrouve certains mots français

Points à améliorer :

- L'enchaînement des mots ne construit pas de phrases logiques. (pas d'idées exprimées)
- Tous les mots utilisés ne sont pas français.

Conclusion pour cet exemple :

L'objectif n'est pas rempli mais la séquence générée présente des motifs de satisfaction. Un entraînement du réseau beaucoup plus long ou avec une puissance de calcul plus grande permettrait un bien meilleur résultat.

Nous avons appris que :

- Entraîner un réseau RNN est long
- Un réseau RNN nécessite une graine sur laquelle s'appuyer pour débiter une séquence.

Exemple 4 :

Objectif :

Agrandir un dictionnaire de mots de passe.

Principe :

On nourrit le réseau avec de nombreux mots de passes courants.

La structure de ces mots de passes courants est donc comprise par le réseau.

On génère de nouveaux mots de passes ne faisant pas partie de la base mais qui présentent les mêmes caractéristiques.

Caractéristiques du réseau de neurone :

Type :

On utilise un réseau de neurone de type RNN LSTM (identique à la génération de poème)

Dimension :

3 couches de taille 128

Fonction d'activation :

Softmax

Graine :

On choisit arbitrairement 3 lettres par lesquels la série de mots de passes générés doit débiter.

Entrées :



Les retour chariots ont été enlevés car ils étaient interprétés comme des caractères (dont la fréquence était très élevée) et étaient ensuite générés à tort.



Pour conserver la cohérence de chaque mot de passe, la base de donnée n'est constituée que de mots de passe de 8 lettres mis bout à bout et le réseau est nourrit par paquet de 8 lettres.

Ainsi, chaque entrée garde sa cohérence et le problème des retours chariots est réglé.

Programme : Voir programme 3

Résultats :

Graine= « mik »

Au bout de 15 minutes :

TEST:
Température =1 :
mikewood
mikistia
mikistin
mikemikm
miklisti
mikmonsu
mikemitc

Température =0.5 :
mikeming
mikemite
mikemitl
mikeming
mikmanch
mikistio
mikiston
mikewood

Au bout de 3h :

Graine : sup, 123, mic

TEST:
Température =1 :
supetouc
superspl
superlas
supethum
supethum
superigi
supethie

Température =0.5 :
supetouc
super1do
supetouc
superlan
superigi
supetouc
super123
supethum

TEST:
Température =1 :
123hardt
123oInkr
123buckd
123555mm
1235esch
123parle
123456te

Température =0.5 :
1234wolf
12323232
1234hono
123repat
123nettt
123repat
123nensu
123555mm

TEST:
Température =1 :
michalls
michmaci
michmacy
michael2
michille
michings
michmong

Température =0.5 :
michillm
michalli
michmone
michandr
michmaca
michillb
michingk
michmons

Points de satisfaction :

- La graine est souvent complétée pour devenir un prénom.
- Les chiffres ne sont pas oubliés et souvent placés de manière cohérentes (en fin, ou de manière à créer une date.
- Certains mots de passes très cohérents (super123, 1234wolf, super1do) sont générés.
- Le principe de graine permet de faire de l'ingénierie sociale.

Points à améliorer :

- Le programme ne sait travailler que sur des mots de passes de longueur identique.

Conclusion pour cet exemple :

L'agrandissement du dictionnaire est réussi mais le renouvellement de la graine à utiliser et la contrainte de taille des mots de passes (il faut relancer l'algorithme pour chaque taille différente) compliquent le processus.

Conclusion

La partie théorique nous a permis de nous familiariser avec les mécanismes utilisés par les réseaux de neurones. Elle était nécessaire pour appréhender les exemples pratiques que nous avons programmé (Notions de dimension du réseau, de fonction d'activation, de graine, etc.). A l'inverse, les exemples pratiques nous ont permis de cerner les contraintes d'utilisation et les limites de ces réseaux. (Base de donnée nécessaire, Puissance/Temps de calcul nécessaire, graine à choisir soi-même pour la génération)

Plus largement, nos recherches nous ont amené à croiser toutes sortes d'utilisation des réseaux de neurones et leurs usages semblent s'appliquer à de nombreux domaines. Reconnaissance d'image, de voix, génération de partition de musique, génération de dialogue.

L'incroyable potentiel de ces réseaux ainsi que la progression continue dans la compréhension et l'utilisation des réseaux de neurones nous ont permis de travailler sur ce sujet avec beaucoup d'enthousiasme du début de nos recherches jusqu'à la fin de la rédaction de ce dossier.