POLYTECHNIQUE MONTRÉAL

LE GÉNIE EN PREMIÈRE CLASSE

INF8225 – I.A. tech. probabilistes et d'apprentissage

Hiver 2020

TP No. 1

Groupe 01

Soumis à : Roger Girgis

7 février 2020

Partie 1

1.1 – Description des taches

Afin de résoudre cette partie, il est nécessaire d'utiliser les quatre concepts suivants :

$$P(A,B) = P(A|B)P(B)\ (Règle\ du\ produit)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}\ (Règle\ de\ Bayes)$$

$$P(X_1) = \sum_{\{X\}\setminus X_1} P(X_1, X_2, \dots, X_N)\ (Règle\ de\ la\ somme)$$

$$P(A|B) = \frac{P(A,B)}{P(B)}\ (Conditioning)$$

a) Pr(H=1) = 0.272
   Dans ce cas, le gazon de Holmes dépend de la variable pluie (P) et de l'arroseur (A).
   Avec la règle de la somme, on obtient donc la probabilité conjointe suivante :
   $$\mathrm{P}(H = 1) = \sum_a \sum_p P(H = 1, P = p, A = a)$$
b) Pr(H=1|W=1) = 0.5956
   En utilisant la règle de Bayes et le concept du conditionnement, on obtient
   $$\mathrm{P}(H = 1|W = 1) = \frac{P(W = 1|H = 1)P(H = 1)}{P(W = 1)} = \frac{\mathrm{P}(H = 1, W = 1)}{P(W = 1)}$$
   Or, nous avons déjà le code pour P(W=1). Ce n'est donc pas une inconnue. Il reste juste à trouver le numérateur. On a que Holmes et Watson dépendent de la variable pluie (P) et de l'arroseur (A). Par la règle de la somme, on a que $\mathrm{Pr}(H = 1, W = 1) = \sum_a \sum_p \mathrm{Pr}(H = 1, W = 1, P = p, A = a)$.
c) Pr(H=1|W=0) = 0.09
   Pour la question c, on utilise le même résonnement que pour le b. On a donc l'équation suivante.
   $$\mathrm{P}(H = 1|W = 0) = \frac{P(W = 0|H = 1)P(H = 1)}{P(W = 0)} = \frac{\mathrm{P}(H = 1, W = 0)}{P(W = 0)}$$
   Pour le numérateur, on utilise la même logique que pour le b. Pour le dénominateur, vu que l'on a $P(W = 1)$, on obtient $P(W = 0)$ avec $P(W = 0) = 1 - P(W = 1)$.
d) Pr(H=1|P=0, W=1) = 0.09
   En utilisant la règle de Bayes, on obtient
   $$\mathrm{P}(H = 1|\mathrm{P} = 0, \mathrm{W} = 1) = \frac{P(\mathrm{P} = 0, \mathrm{W} = 1|H = 1)P(H = 1)}{P(\mathrm{P} = 0, \mathrm{W} = 1)}$$
   En utilisant le concept de conditionnement, on obtient
   $$\frac{P(\mathrm{P} = 0, \mathrm{W} = 1|H = 1)P(H = 1)}{P(\mathrm{P} = 0, \mathrm{W} = 1)} = \frac{P(\mathrm{P} = 0, \mathrm{W} = 1, H = 1)}{P(\mathrm{P} = 0, \mathrm{W} = 1)}$$

Ensuite, il faut appliquer la règle de la somme afin de résoudre la probabilité.

e) Pr(W=1|H=1)

Cette question utilise la même logique que le b. En utilisant la règle de Bayes et le concept du conditionnement, on obtient

$$\mathrm{P}(W=1|H=1)=\frac{P(H=1|W=1)P(W=1)}{P(H=1)}=\frac{\mathrm{P}(H=1,W=1)}{P(H=1)}$$

On connait $P(H=1)$ à l'aide du a.

f) Pr(W=1|H=1, A=1)

En utilisant la règle de Bayes, on obtient

$$\mathrm{P}(W=1|H=1,A=1)=\frac{P(H=1,A=1|W=1)P(W=1)}{P(H=1,A=1)}$$

En utilisant le concept de conditionnement, on obtient

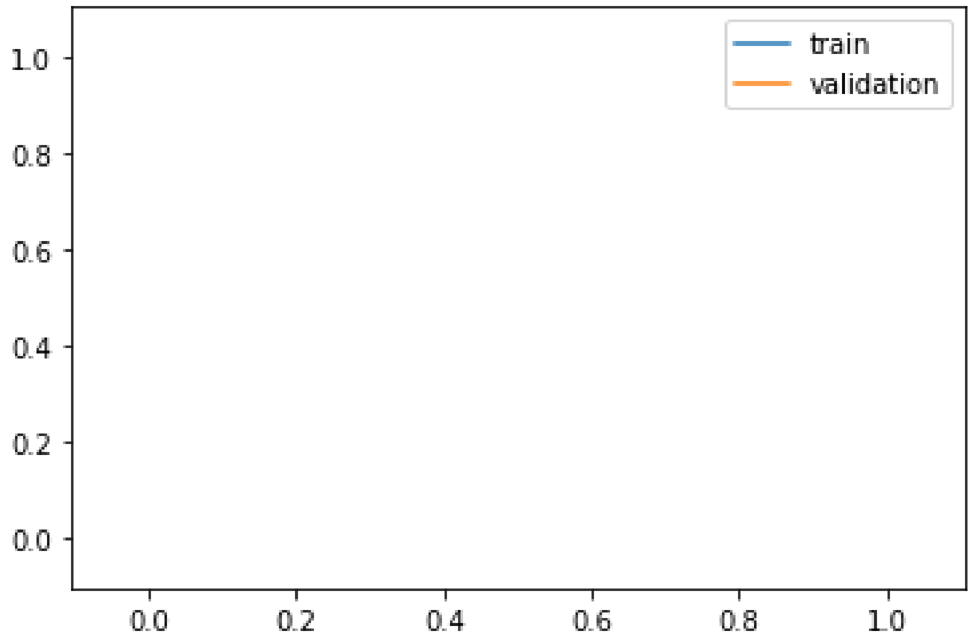
$$\frac{P(H=1,A=1|W=1)P(W=1)}{P(H=1,A=1)}=\frac{P(H=1,A=1,W=1)}{P(H=1,A=1)}$$

Ensuite, il faut appliquer la règle de la somme afin de résoudre la probabilité.

Partie 2

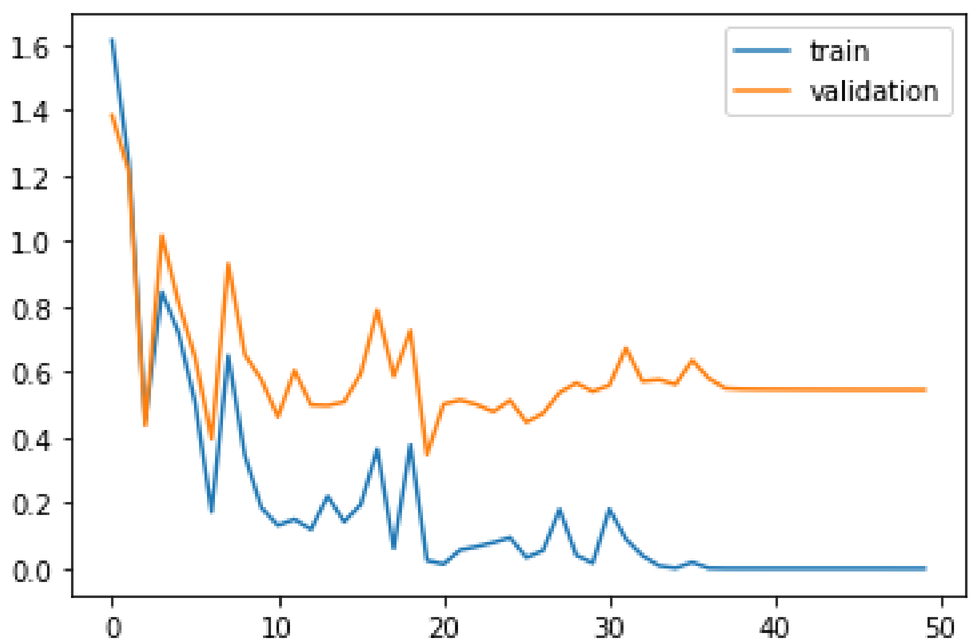
Question A

Figure1- Courbe de log de vraisemblance pour un lr=0.1 et un mini batch size=1



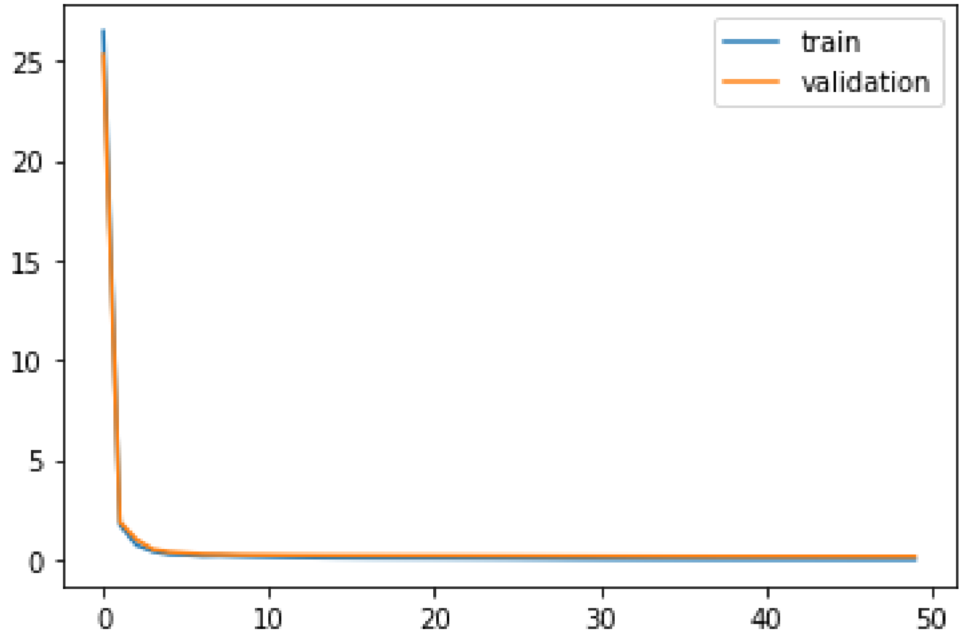
Accuracy: 0.9666666666666667

Figure2- Courbe de log de vraisemblance pour un lr=0.1 et un mini batch size=20



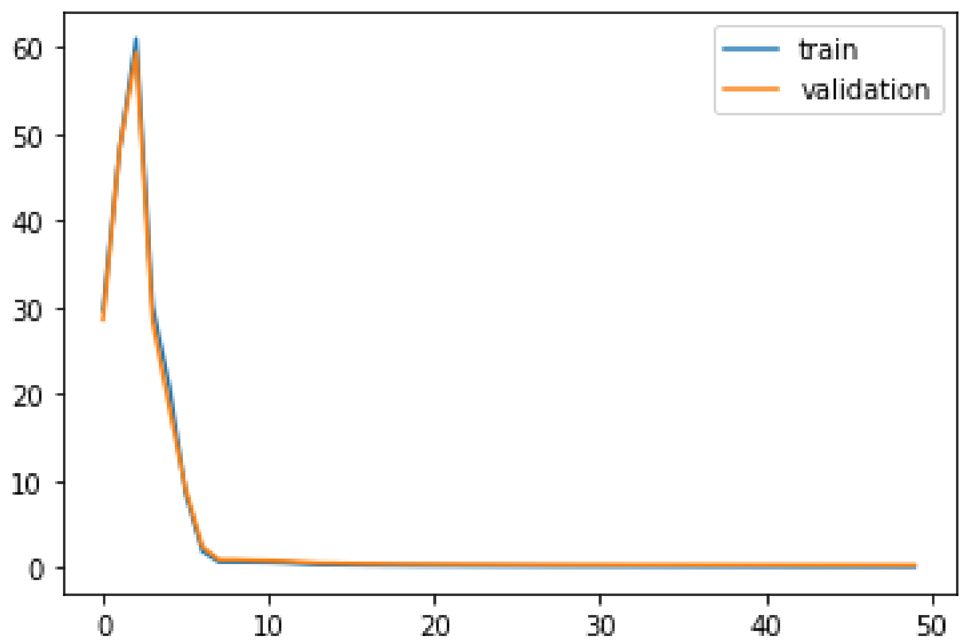
Accuracy: 0.9666666666666667

Figure3- Courbe de log de vraisemblance pour un lr=0.1 et un mini batch size=200



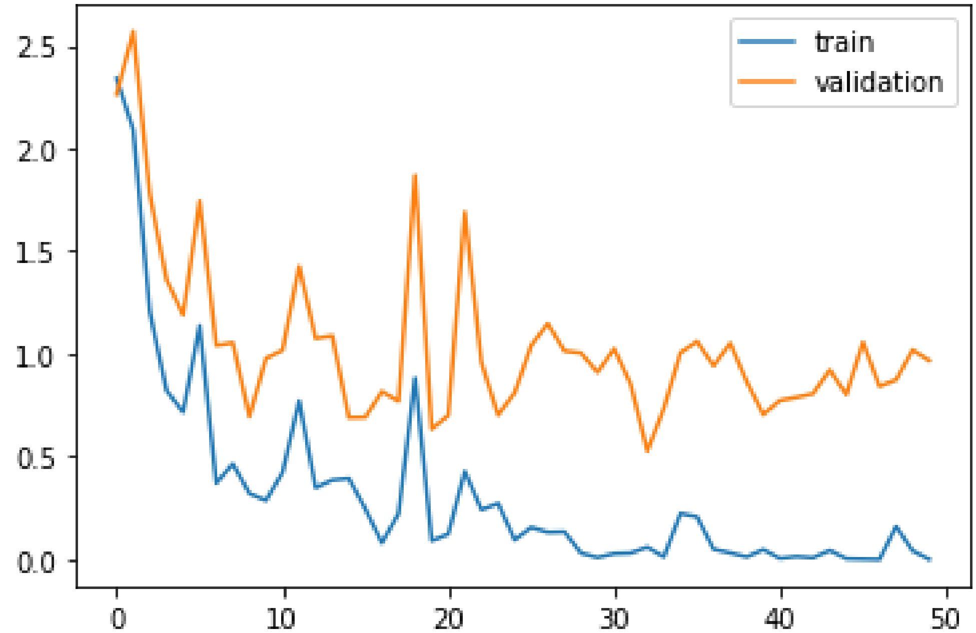
Accuracy: 0.9629629629629629

Figure4- Courbe de log de vraisemblance pour un lr=0.1 et un mini batch size=1000



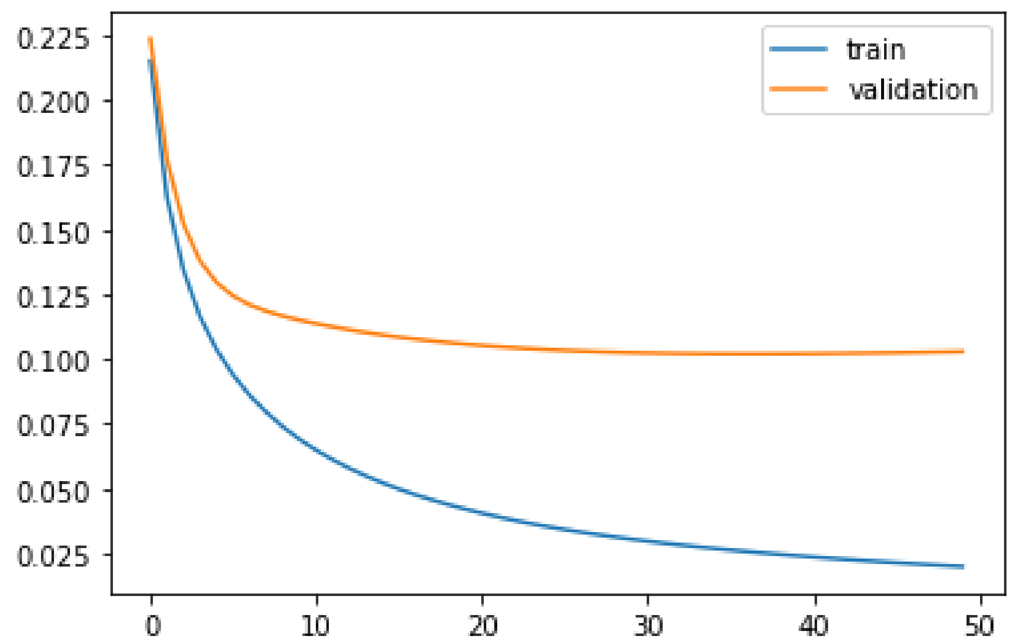
Accuracy: 0.9592592592592593

Figure5- Courbe de log de vraisemblance pour un lr=0.01 et un mini batch size=1



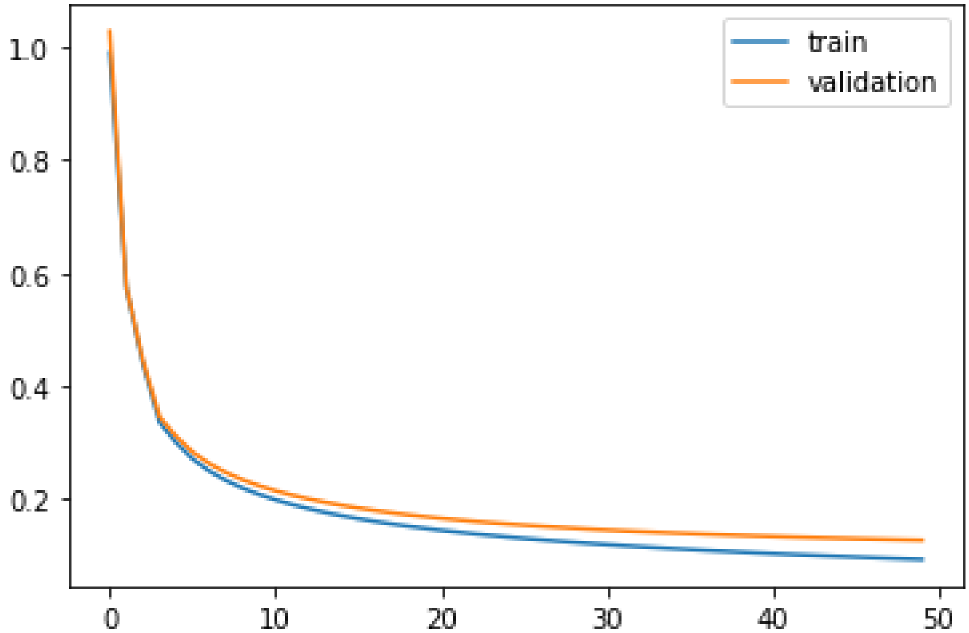
Accuracy: 0.9555555555555556

Figure6- Courbe de log de vraisemblance pour un lr=0.01 et un mini batch size=20



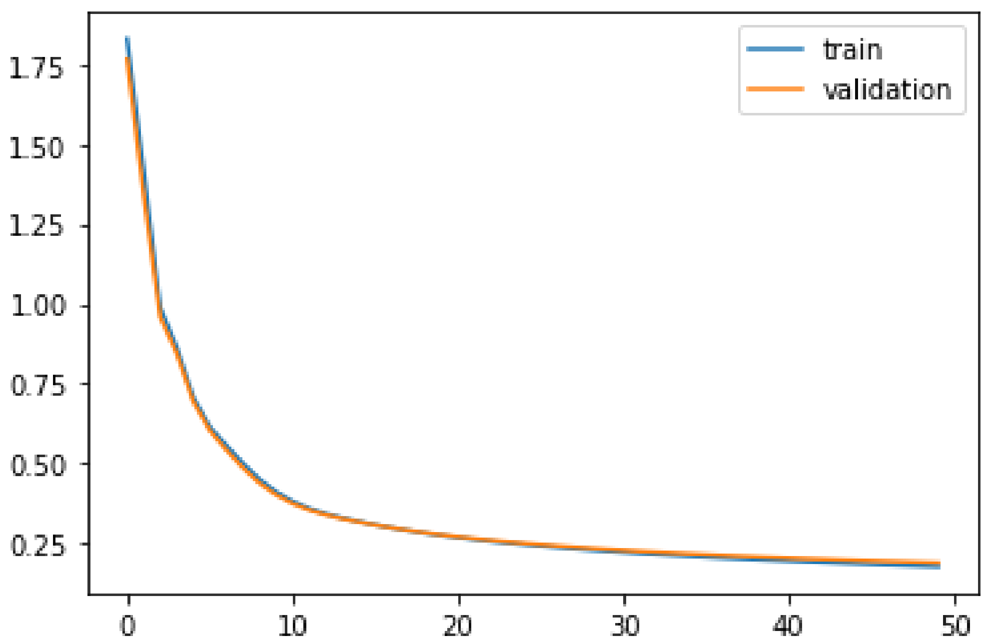
Accuracy: 0.9666666666666667

Figure7- Courbe de log de vraisemblance pour un lr=0.01 et un mini batch size=200



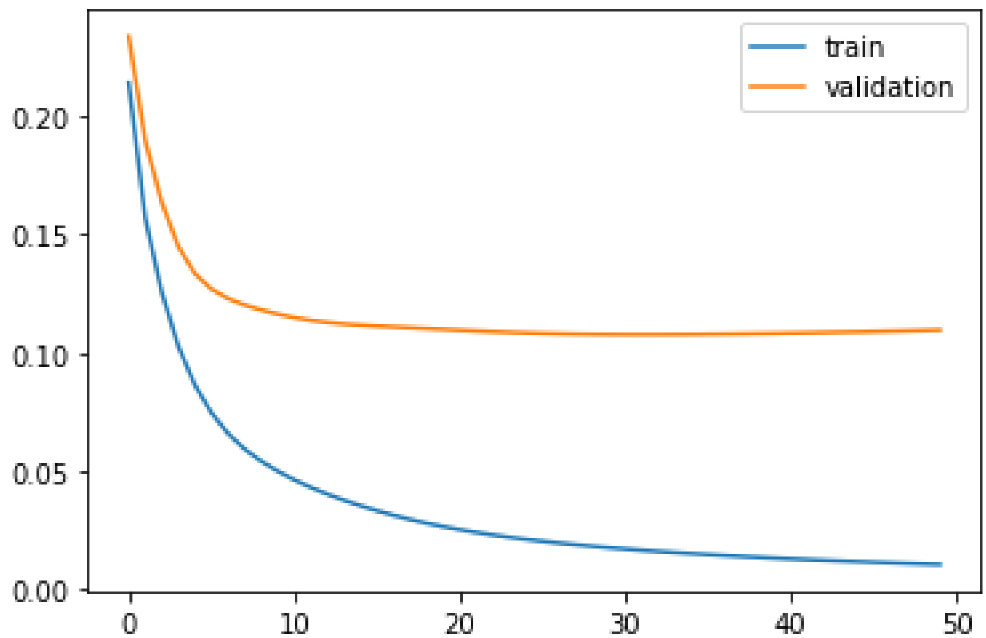
Accuracy: 0.9629629629629629

Figure8- Courbe de log de vraisemblance pour un lr=0.01 et un mini batch size=1000



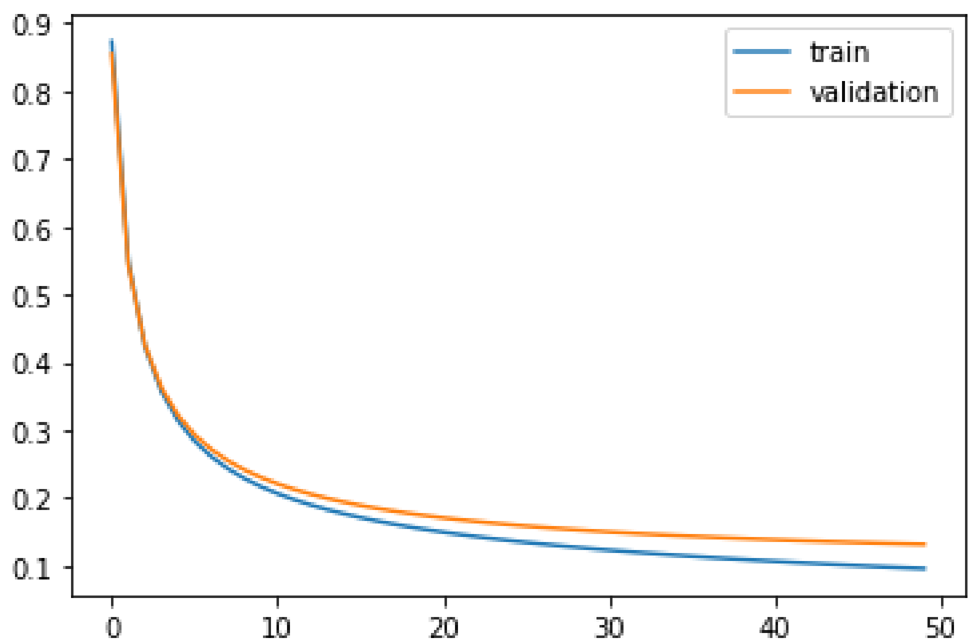
Accuracy: 0.9518518518518518

Figure9- Courbe de log de vraisemblance pour un lr=0.001 et un mini batch size=1



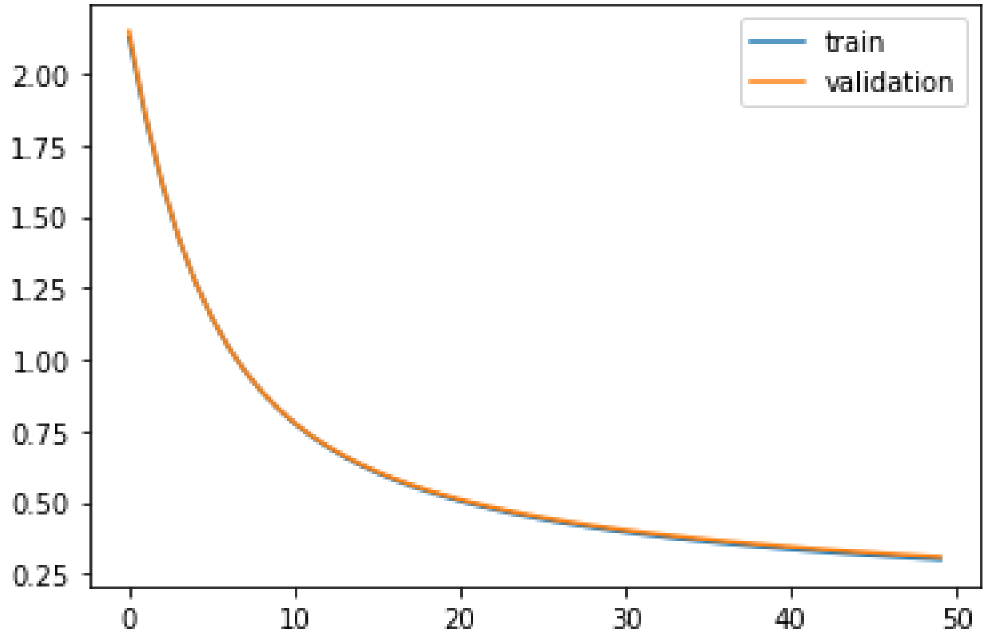
Accuracy: 0.9703703703703703

Figure10- Courbe de log de vraisemblance pour un lr=0.001 et un mini batch size=20



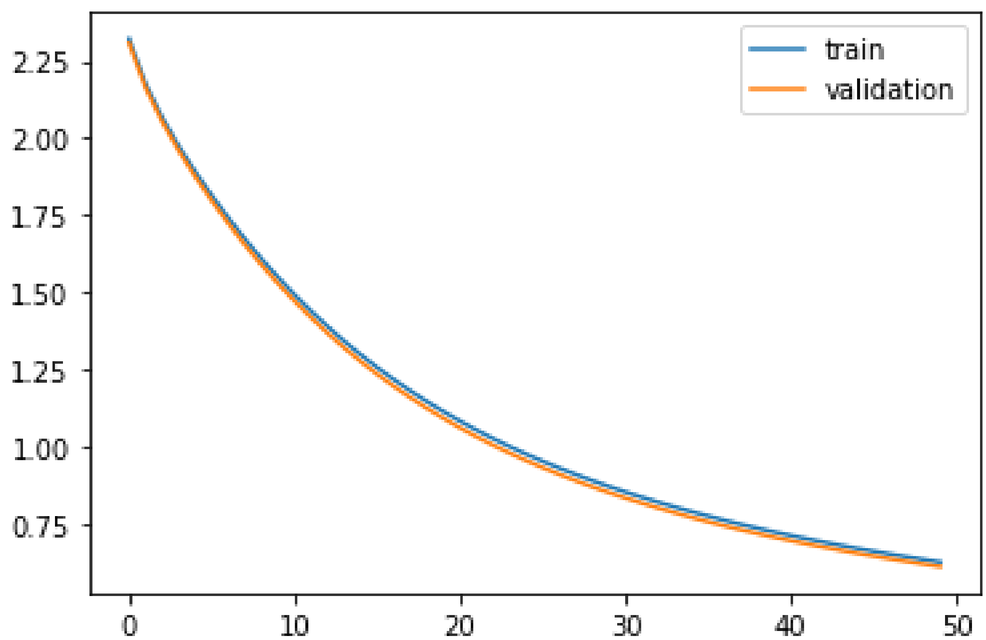
Accuracy: 0.9666666666666667

Figure11- Courbe de log de vraisemblance pour un lr=0.001 et un mini batch size=200



Accuracy: 0.9296296296296296

Figure12- Courbe de log de vraisemblance pour un lr=0.001 et un mini batch size=1000



Accuracy: 0.9185185185185185

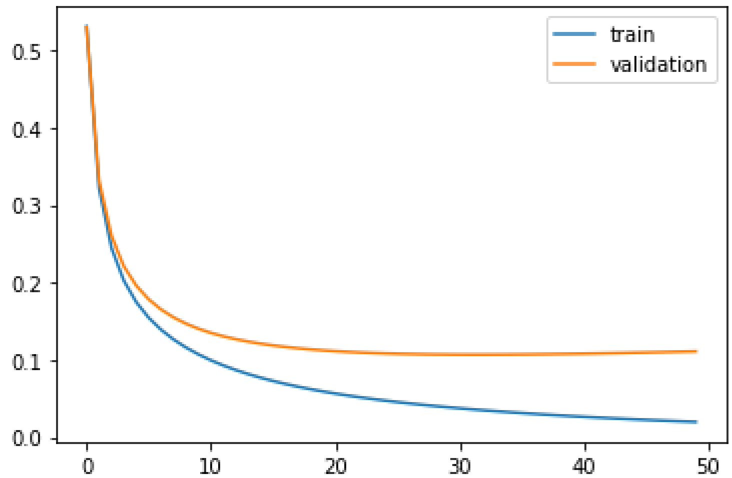
Comme on peut le voir, la figure 1 est vide. En effet, cela est dû au fait que notre fonction *get_loss()* doit faire un logarithme de 0. C'est ce qui cause que notre liste *losses* contient des NaN (*not a number*).

En regardant les 12 figures, on regarde qu'il n'y a pas de surapprentissage qui se fait.

Au vu des résultats, les paramètres donnant le meilleur résultat, le learning rate est de 0.001 et la taille des mini-batch est de 20. En effet, bien qu'il a un plus petit accuracy que la figure 9, la valeur de loss est plus petite.

Question B

```
Figure 13 - Courbe de log de vraisemblance
```




```
Accuracy: 0.9555555555555556
```

À partir de la figure 13, on remarque qu'il y a surapprentissage. En effet, à partir de l'époque 35, la courbe de validation augmente de plus en plus. Si l'on compare avec la figure 10, on remarque que le loss pour les valeurs de validation n'a pas vraiment changé. Le loss pour les valeurs d'entrainement ont diminué, mais cela est surement dû au surapprentissage. De plus, on remarque que l'accuracy a diminué.