

FACULTÉ DES SCIENCES ET INGÉNIERIE

SORBONNE UNIVERSITÉ

RECONNAISSANCE DES FORMES POUR L'ANALYSE ET
L'INTERPRÉTATION D'IMAGES

ADVANCED DEEP LEARNING FOR IMAGE UNDERSTANDING

Rapport de TMEs 12 à 14

Auteurs :

Bochra DOUIDI

Ahmed Tidiane BALDE

Encadrants :

Nicolas THOME

Charles CORBIERE

29 janvier 2020



TABLE DES MATIÈRES

Bayesian Models and Deep Learning	2
TP1 : Bayesian Linear Regression	2
Part I : Linear Basis function model	2
Q1.2	2
Q1.5	3
Bonus Question	3
Part II : Non Linear models	4
II.1 Polynomial basis functions	4
Q2.2	4
II.2 Gaussian basis functions	4
Q2.4	4
Q2.5	4
TP 2 : Approximate Variational Inference	5
Part I : MC Dropout variational inference in regression	5
Q1.2	5
Q1.3	5
Part II : Bayesian Logistic Regression	6
II.1 First solution : MAP estimate	6
Q2.3	6
II.2 Second solution : Variational Inference	6
Q2.5	6
II.3 MLP with MCDropout variational inference	7
Q2.8	7
TP 3 : Uncertainty Applications	7
Part I : MC Dropout on MNIST	7
Q1.3	7
Q1.4	7
Part II : Failure prediction	9
Q2.3	9
Part III : Out-of-distribution detection	9
Q3.1	9
Références	10

BAYESIAN MODELS AND DEEP LEARNING

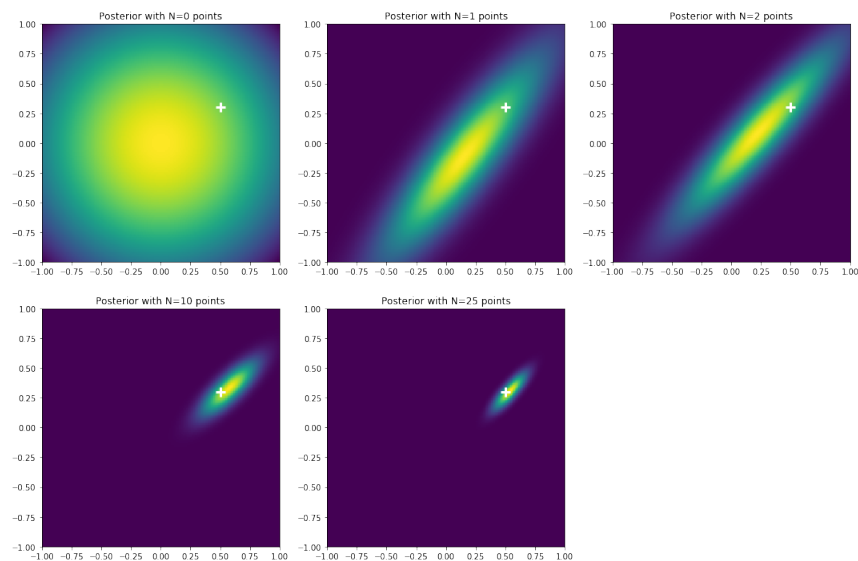
Avec l'essor de la *vision par ordinateur*, les scientifiques tout comme les amateurs se sont vus confrontés à de nouveaux problèmes, dont certains étant la fiabilité que nous pouvons accorder à ces systèmes. Nous avons dans les cours qui précéderent celui-ci, vu les *réseaux adversaires*, consistant à des techniques pour tromper un réseau de neurones. Nous étions stupéfaits de la facilité à y parvenir. C'est un problème d'une grande ampleur qui, par ailleurs, justifie la réticence de certains acteurs à leur utilisation, notamment dans les domaines tels que l'automobile ou la médecine, où le moindre faux positif pourrait résulter à des conséquences désastreuses.

Notre but n'est pas de faire des réseaux parfaits qui ne se trompent pas, se serait d'ailleurs très suspect, quand bien même nous le voulons. L'objectif est d'avoir un réseau qui, dans une situation délicate, pourrait alors nous faire part de son incertitude en nous disant qu'il ne sait pas. Plutôt que de prendre une décision coûte que coûte. C'est alors qu'intervient les *réseaux bayésiens*.

TP1 : BAYESIAN LINEAR REGRESSION

PART I : LINEAR BASIS FUNCTION MODEL

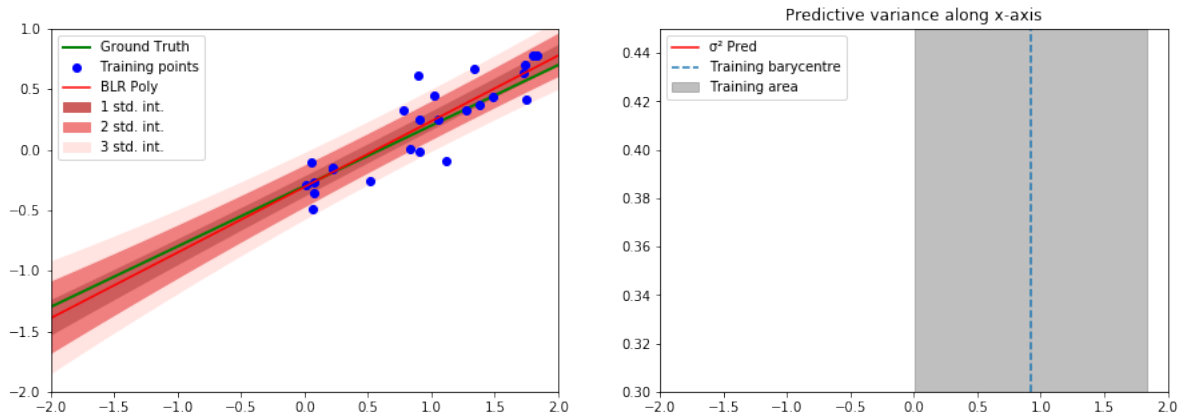
Q1.2



Ici nous expérimentons le *posterior sampling*, nous pouvons observer notamment que quand le nombre de points est de 0, le *posterior* est égal au *prior* et plus le nombre de points augmente, plus on approche la cible, la croix blanche.

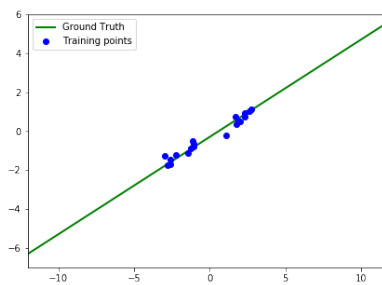
Q1.5

Plus nous nous éloignons des points, plus la variance prédictive augmente. Cela est dû naturellement au fait que nous n'avons pas d'informations quand nous nous éloignons de la représentation de nos données. En somme, il suffit de se placer dans un graphe 2D où nous avons un seul point, bien entendu, nous aurons alors une variance très élevée, et plus nous aurons des points, plus la variance sera faible. Si nous avons tous les points de l'espace, nous aurons alors une variance nulle.

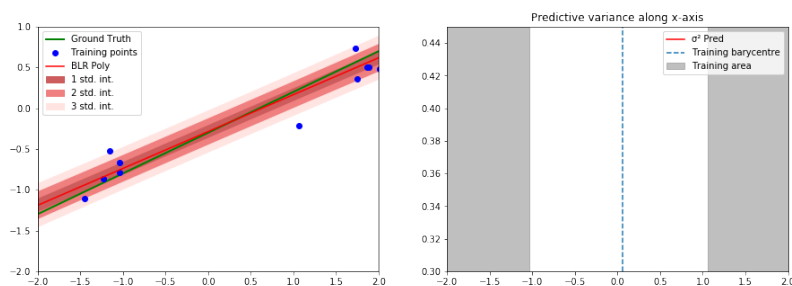


BONUS QUESTION

Dans cet exemple, à première vue, nous voyons toute suite que les données sont très corrélées, et la variance est très faible parce qu'en traçant une régression linéaire on pourrait presque les modéliser à la perfection. Alors l'écart-type est constant à ce niveau.



Bonus data points



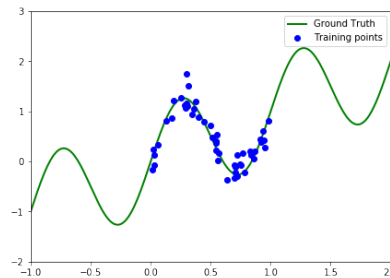
Bayesian linear regression

PART II : NON LINEAR MODELS

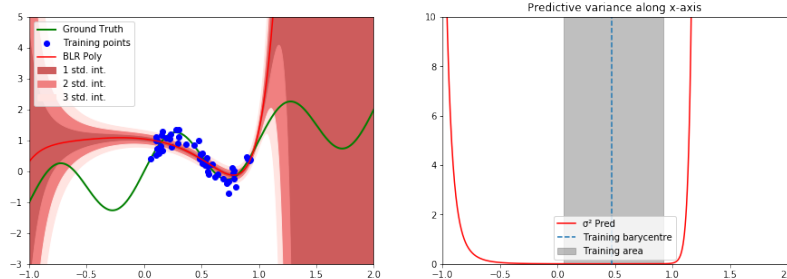
II.1 POLYNOMIAL BASIS FUNCTIONS

Q2.2

Maintenant, faisons l'expérience sur des données non linéaires. Notre régression précédente étant linéaire, il faut donc l'adapter pour pouvoir modéliser nos nouvelles données. C'est pourquoi nous les projetons dans de nouvelles dimensions à l'aide des kernels, en l'occurrence, on utilise le kernel *polynomial*.



Sinus data points



Bayesian regression with polynomial kernel

Alors quand on s'éloigne des données, la variance prédictive augmente drastiquement. Au tout début, elle suit la forme de nos données sinusoïdales, et s'élargit sur toutes la zone par après. Nous expliquons cela, par la forme de nos données, elles ne sont plus linéaires, et peuvent éventuellement être dispersées. Nous pensons alors qu'il est très juste d'avoir une variance qui évolue ainsi.

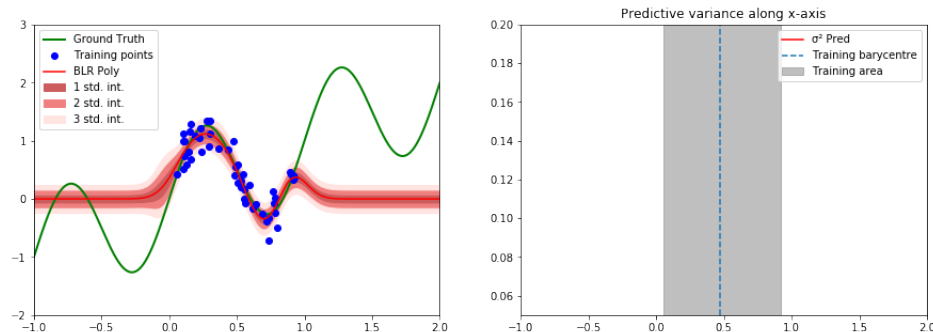
II.2 GAUSSIAN BASIS FUNCTIONS

Q2.4

Le Kernel *bayésien*, un peu plus complexe que celui *polynomial*, modélise mieux nos données.

Q2.5

La variance prédictive converge vers 0 au fur et à mesure que nous nous éloignons des données. Cela doit probablement être dû au fait que la prédiction avec le kernel gaussien est distribué autour de la moyenne.



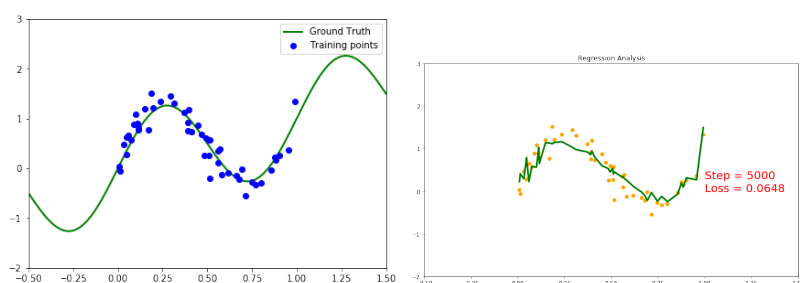
Bayesian regression with gaussian kernel

TP 2 : APPROXIMATE VARIATIONAL INFERENCE

PART I : MC DROPOUT VARIATIONAL INFERENCE IN REGRESSION

Q1.2

Les réseaux de neurones bayésiens utilisent le même principe que celui vu précédemment. À la place des poids, nous avons maintenant des gaussiennes avec leur moyenne comme prédiction y .

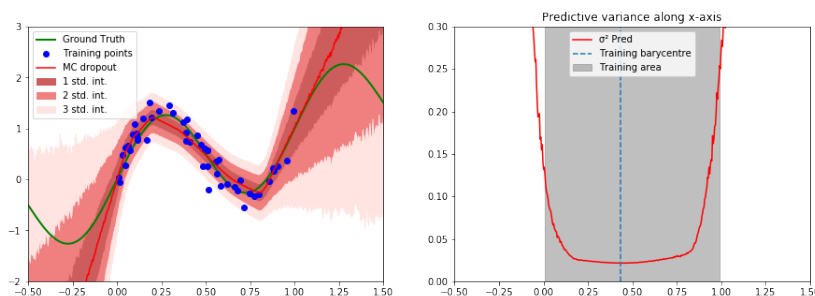


Data points

Regression

Q1.3

En activant le *dropout* en test avec une probabilité d'éteindre certaines neurones définies à l'avance ainsi que le nombre de fois où la prédiction est effectuée en l'occurrence 1000, nous tâchons d'une certaine manière d'impliquer tous les neurones dans la prédiction. Nous pouvons dès lors observer que la variance prédictive évolue d'une manière assez pertinente en apprenant la forme de nos données.

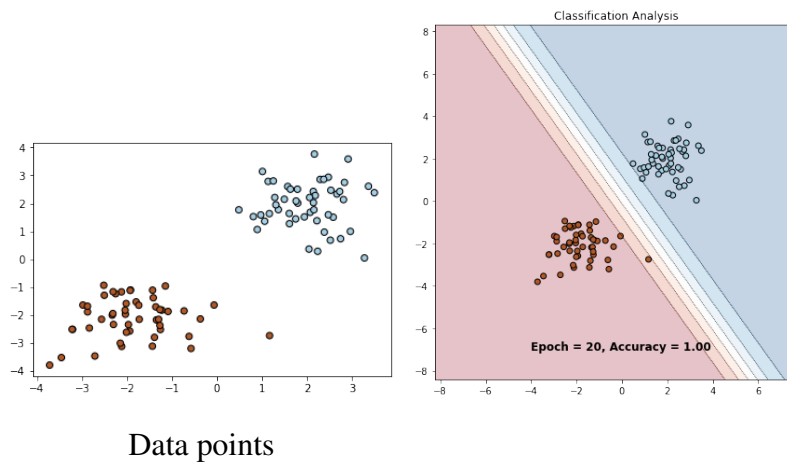


PART II : BAYESIAN LOGISTIC REGRESSION

II.1 FIRST SOLUTION : MAP ESTIMATE

Q2.3

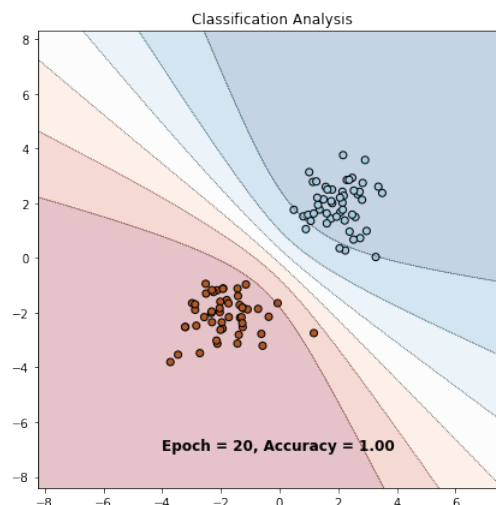
Le principal désavantage de cette méthode est que nous ne pouvons obtenir une distribution sur w . Les points éloignés de nos données d'apprentissage seront alors classés suivant l'hyperplan obtenu.



II.2 SECOND SOLUTION : VARIATIONAL INFERENCE

Q2.5

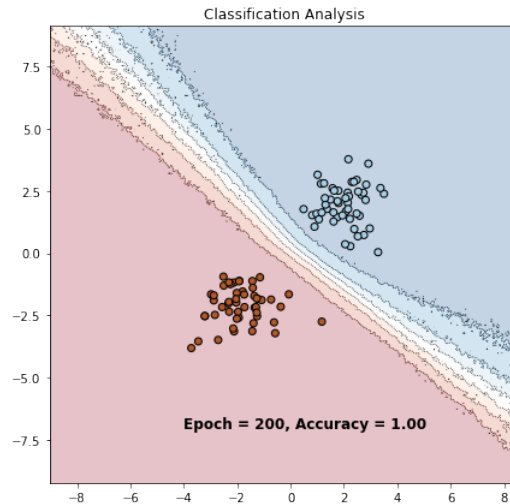
Le but ici est de trouver la distribution postérieure. Cela n'étant pas possible, nous optons donc pour une approximation de cette distribution. Ainsi, nous observons qu'avec cette méthode, plus on s'éloigne des données encore une fois, plus la variance augmente. Les points non proches de notre distribution de données seront donc classifiés en conséquence.



II.3 MLP WITH MCDROPOUT VARIATIONAL INFERENCE

Q2.8

En utilisant le *dropout* comme approximation bayésienne (Gal and Ghahramani, 2015). Nous retrouvons des résultats similaires aux précédents, mais légèrement meilleurs. La variance prédictive est à notre goût beaucoup plus prudente et juste vis à vis des données d'apprentissage.



TP 3 : UNCERTAINTY APPLICATIONS

Dans cette partie, nous avons vu plusieurs méthodes pour estimer l'incertitude d'un modèle dont certaines en utilisant la fréquence des classes prédites, d'autres l'entropie ou encore l'information mutuelle.

PART I : MC DROPOUT ON MNIST

Après avoir entraîné notre réseau de neurones *LeNet* sur 20 epochs, nous l'avons sauvegardé pour réaliser les expériences ci-après.

Q1.3

Ci-dessous, les quatre (4) exemples, dont les prédictions étaient les plus incertaines. La méthode *var-ratios*, se sert simplement de l'histogramme des fréquences des classes prédites pour chaque exemple et calcule dès lors l'incertitude. Nous pouvons voir à titre d'exemple, le 2 qui est confondu à 7 et ainsi de suite.

Q1.4

En comparant les résultats obtenus ci-haut à ceux obtenus avec des exemples pris au hasard. Ne représentant donc pas un quelconque ordre dans le top des exemples incertains, nous remarquons très rapidement la différence. Naturellement pour ces chiffres, il n'y a pratiquement pas d'ambiguïtés quant à leur classe respective.

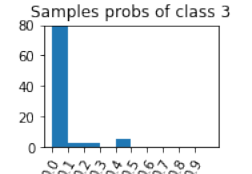
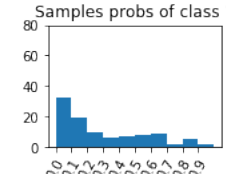
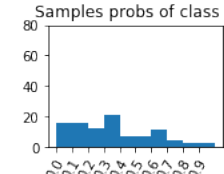
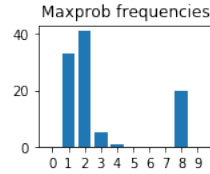
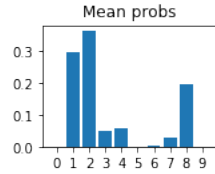
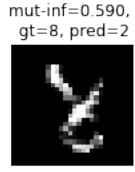
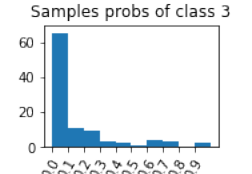
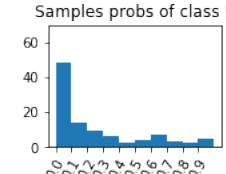
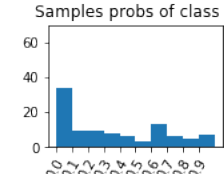
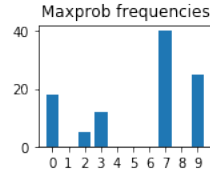
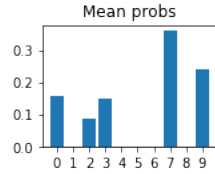
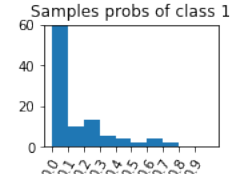
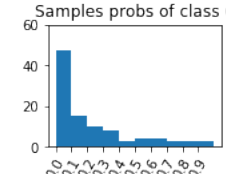
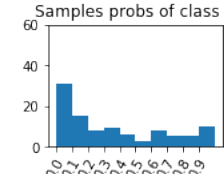
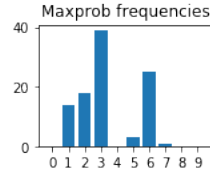
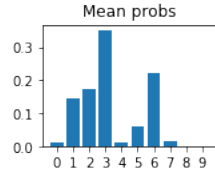
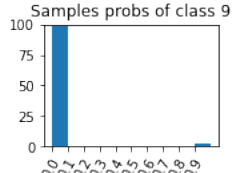
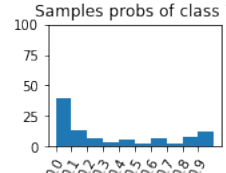
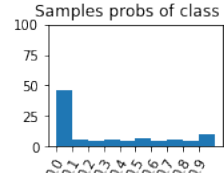
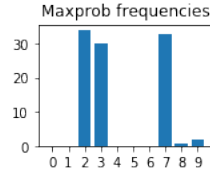
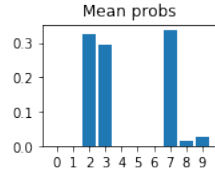
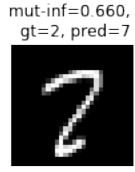


FIGURE 1 – Q1.3, Top4_uncertain_samples

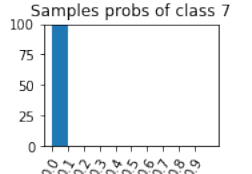
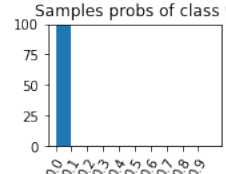
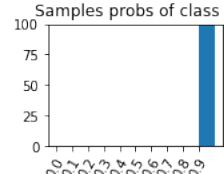
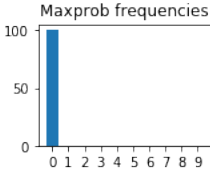
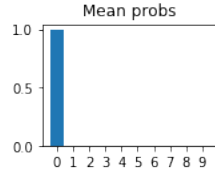
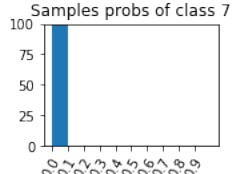
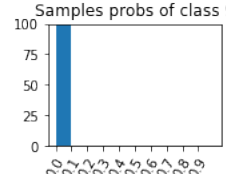
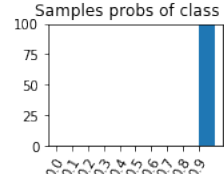
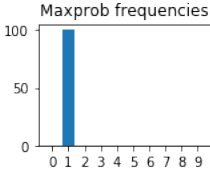
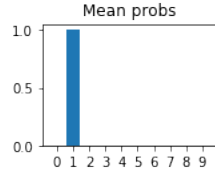
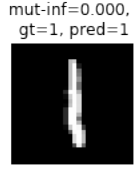
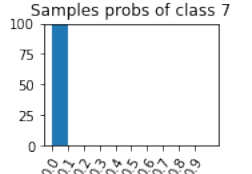
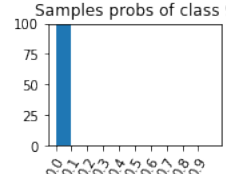
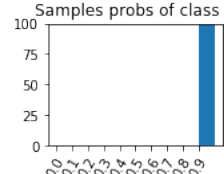
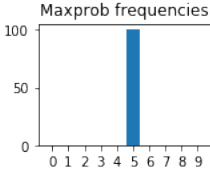
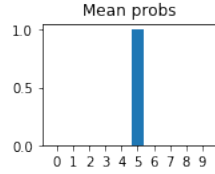
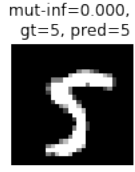
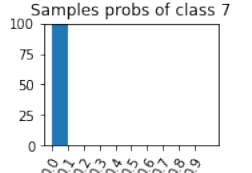
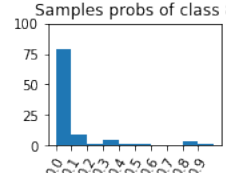
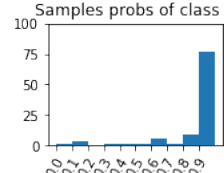
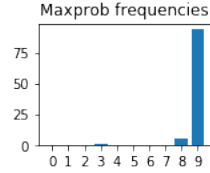
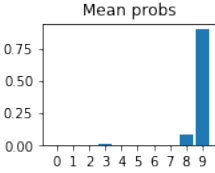
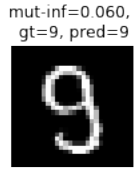
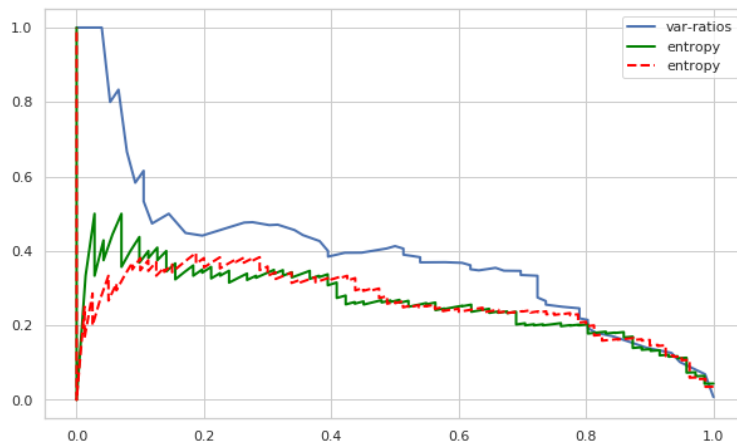


FIGURE 2 – Q1.4, 4_random_uncertain_samples

PART II : FAILURE PREDICTION

Q2.3

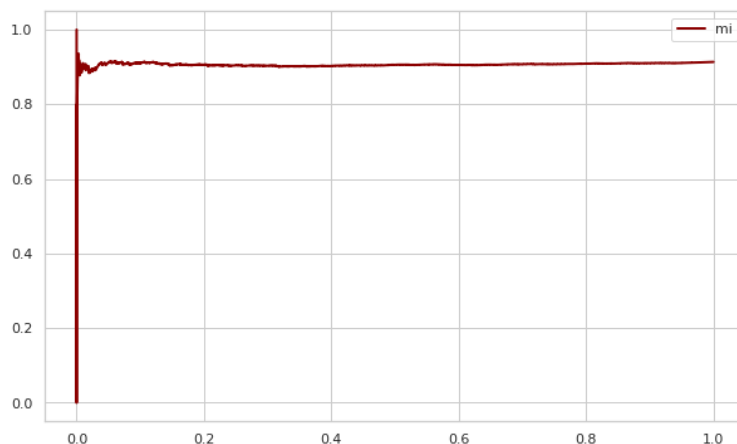
Rappelons que la précision mesure le taux de faux positifs et le rappel, le taux de faux négatifs, qui sont par ailleurs plus dangereux quand on a affaire avec des problèmes délicats. Ci-dessous, les trois courbes *AUPR* pour les trois différentes méthodes d'estimations d'incertitudes que nous avons vues. Les résultats de la méthode utilisant l'information mutuelle sont assez similaires à ceux de l'entropie. Très certainement dû fait que la première utilise la seconde. Ces deux méthodes montrent ici l'incertitude autour des faux négatifs car le rappel est nul pour de nombreux exemples sauf quand la précision est aux alentours de 0.4, 0.2. Quant à la méthode *var-ratios*, nous remarquons que le rappel et la précision évolue presque simultanément.



PART III : OUT-OF-DISTRIBUTION DETECTION

Q3.1

Ici, le principe de *out – of – distribution* se fait sentir dans les résultats. Quand le rappel est très fort, la précision l'est également, ce qui généralement veut dire que le modèle est pire qu'un modèle random.



RÉFÉRENCES

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation : Representing model uncertainty in deep learning, 2015.