

Classification des phases du modèle d'Ising

Coraline Letouzé

4 janvier 2021

1 Contexte

Il s'agit ici de mettre en application les outils du machine learning (ML) sur un problème de physique, en l'occurrence identifier les phases ordonnée et désordonnée du modèle d'Ising. On cherchera également comparer les résultats obtenus entre le modèle d'Ising habituel aux interactions ferromagnétiques et sa version anti-ferromagnétique. On utilisera pour cela une analyse en composantes principales suivie d'une régression logistique, et un réseau de neurones convolutif.

Le modèle d'Ising 2D Le modèle d'Ising à 2 dimensions est un modèle de spins sur réseau carré, avec des interactions entre premiers voisins. On considère uniquement les composantes des spins selon l'axe z , qui prennent leur valeur dans $\{+, -\}$. Son hamiltonien en champ magnétique nul est ainsi donné par (Éq. 1).

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j \quad (1)$$

où σ_i est le moment magnétique de spin au nœud i . La sommation est menée sur les couples $\langle i, j \rangle$ de premiers voisins.

J est un terme de couplage effectif entre spins (interaction d'échange), que l'on choisit constant. S'il est positif ($J = +1$), les spins tendent à s'aligner et on observe une phase ordonnée à basse température, ainsi qu'une magnétisation spontanée. On parle de système ferromagnétique. À haute température, les effets entropiques dominent et l'ordonnancement disparaît (Figure 1). La magnétisation est alors proportionnelle au champ magnétique (paramagnétisme). Entre les deux se produit une transition de phase à la température critique dite de Curie. Cette température vaut $T_c \approx 2.27$ dans le système d'unité choisi ($|J| = 1, k_B = 1$).

Si $J = -1$, la configuration la plus stable à température nulle présente une alternance de spins $+$ et $-$. Le système est ordonné mais contrairement au cas ferromagnétique la magnétisation totale est nulle. On parle de

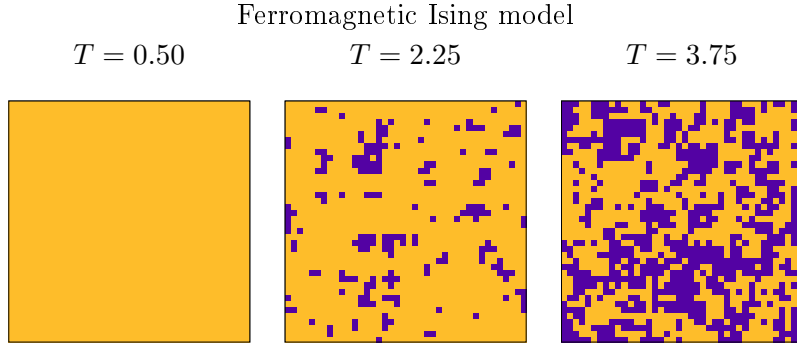


FIGURE 1 – Configurations du modèle d’Ising ferromagnétique selon la température.

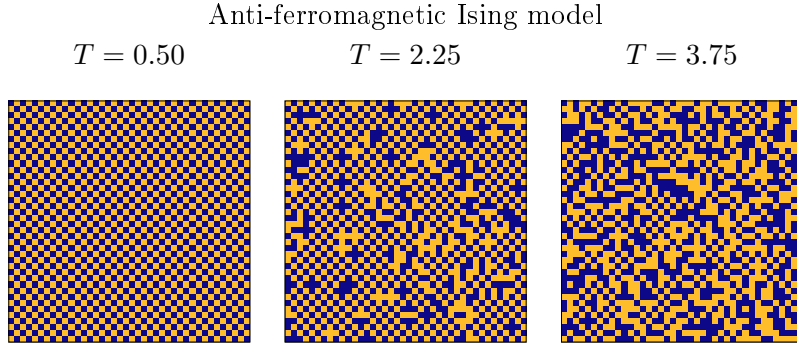


FIGURE 2 – Configurations du modèle d’Ising anti-ferromagnétique selon la température.

comportement anti-ferromagnétique. Là encore le système présente une transition entre la phase ordonnée anti-ferromagnétique et la phase désordonnée paramagnétique (Figure 2).

Les grandeurs physiques intéressantes sont d’une part l’énergie $E = H$ et la magnétisation $M = \sum_i \sigma_i$, qui peuvent être calculées directement à partir de la configuration des spins. D’autre part leurs dérivées, à savoir la capacité calorifique et la susceptibilité magnétique, divergent à la transition de phase, mais sont plus difficiles à calculer puisqu’elles requièrent la connaissance des fluctuations du système.

Ce modèle d’Ising (ferromagnétique) à deux dimension est très utile d’un point de vue théorique car c’est le modèle le plus simple pour décrire la transition de phase para- / ferro- magnétique. En outre sa solution exacte (en champ nul) étant connue, il permet de comparer différentes méthodes d’approximation (par exemple le champ moyen). Il est en outre bien adapté aux

calculs numériques et peut décrire d'autres phénomènes physiques au-delà du ferromagnétisme, comme les alliages à deux constituants ou un modèle simple de gaz sur réseau.

Inspirations Le modèle d'Ising, de par son importance théorique et sa simplicité de calcul, a été extensivement étudié, y compris par les outils du ML. On citera comme source d'inspiration principale l'article [4] pour sa présentation limpide des outils et méthodes du ML, dont les applications au modèle d'Ising sont succinctement illustrées. L'idée d'appliquer une PCA provient de la présentation [2].

Outils Les données utilisées proviennent d'une part de l'ensemble généré pour [4], d'autre part de simulation Monte-Carlo personnelle. Le développement a été fait sur Jupyter notebook et Python, en utilisant les bibliothèques Numpy, Pickle, Matplotlib ; Scikit-Learn a fourni les outils du ML et Keras les méthodes propres aux réseaux de neurones. Les graphes ont été tracés sous Gnuplot et le présent rapport mis en page par L^AT_EX.

2 Les données

Configurations ferromagnétiques On travaille avec les données générées pour la synthèse [4], [5]. Elles consistent en 10000 échantillons par température, qui varie de $T = 0.25$ à $T = 4.0$ par pas de 0.25. Les configurations sont de dimension 40×40 , décorréées, et exploitent la symétrie d'inversion de spin $\sigma_i \rightarrow -\sigma_i$ [5].

Modèle d'Ising antiferromagnétique Les configurations antiferromagnétiques sont générées par une simulation Monte-Carlo utilisant l'algorithme de Metropolis (décrit par exemple dans [3]). Le système est simulé au même format $(40, 40)$ et aux mêmes températures que les configurations ferromagnétiques.

L'algorithme de Metropolis consiste à proposer le retournement d'un spin à chaque itération. Si cette offre est énergétiquement favorable, elle est systématiquement acceptée ; dans le cas contraire, la probabilité qu'elle soit validée est donnée par le facteur de Boltzmann associé à la différence d'énergie. On appellera cycle une séquence de N offres, où N est le nombre de spins dans le réseau (ici $N = 1600$).

On génère ainsi 20 échantillons par température, indépendants. Ils sont prélevés après 250 cycles, suffisant pour passer d'une configuration initiale aléatoire à une configuration ordonnée. Chacun d'eux est ensuite transformé par rotation, miroir et inversion Z_2 ($S \rightarrow -S$). Ces transformations préservent les relations locales entre spins. On obtient ainsi 16 fois plus de données.

Ensembles d’entraînement et de test Suivant le choix présenté dans [4], les ensembles d’entraînement et de validation sont constitués uniquement de configurations aux phases facilement identifiables : jusqu’à $T = 1.75$ pour la phase ordonnée (étiquette 0) (étiquette 1), à partir de $T = 2.75$ pour la phase désordonnée. Les températures intermédiaires, au voisinage de la température critique, forment l’ensemble de test. On dispose donc pour le système ferromagnétique de 130000 échantillons étiquetés, 30000 échantillons non identifiés. Pour les données anti-ferromagnétiques, les ensembles sont bien plus modestes : 4160 échantillons de phase bien déterminée, et 960 près de T_c .

3 Principal Component Analysis (PCA) – Analyse en Composantes Principales

Analyse en composantes principales La PCA est une méthode d’analyse en statistique multivariée qui permet de décorréler et de réduire la dimensionnalité des variables en les dé-bruitant. Pour un ensemble de points dans un espace de dimension p , on appelle composantes principales de cet ensemble une famille orthonormée de p vecteurs. Géométriquement, ces vecteurs forment les axes d’un ellipsoïde de dimension p ajusté aux données. Il s’agit également des vecteurs propres de la matrice de covariance de l’ensemble de points, dont les valeurs propres correspondent à la variance de la projection des points sur l’axe considéré. Ces axes sont appelés composantes principales car intuitivement, plus la variance associée est élevée, plus l’axe est porteur d’information. On projette alors les points sur les $n < p$ premières (dans le sens de la contribution à la variance totale) composantes principales ; les axes ignorés sont considérés bruités ou moins significatifs [7].

Résultat de la PCA sur le système ferromagnétique Concernant le modèle d’Ising ferromagnétique, cette approche est très efficace. En effet la première composante explique à elle seule plus de la moitié de la variance totale ; la deuxième composante contribue à seulement 0.3% (Table 1). L’importance de la première composante est confortée par l’argument physique suivante : les poids associés à la première composante (Figure 3) sont quasiment égaux (variations en 10^{-4}) ; elle correspond donc à peu près à la magnétisation de l’échantillon. Or on sait que la magnétisation est le paramètre d’ordre de la transition para/ferro magnétique, qui s’annule théoriquement dans la phase désordonnée.

Si on regarde la répartition des données selon les deux premières composantes (Figure 4), on constate qu’elles se répartissent en trois zones distinctes. L’amas central correspond aux magnétisations proches de 0, autrement dit aux températures élevées et à la phase désordonnée. Les deux amas à droite et à gauche correspondent à des magnétisations fortes (basses températures

Results	Ferromagnetic		Antiferromagnetic		
	PCA 1	PCA 2	PCA 1	PCA 2	PCA
Explained variance (%)	52.8	0.3	33.3	3.5	3.1

TABLE 1 – Résultats de la PCA.

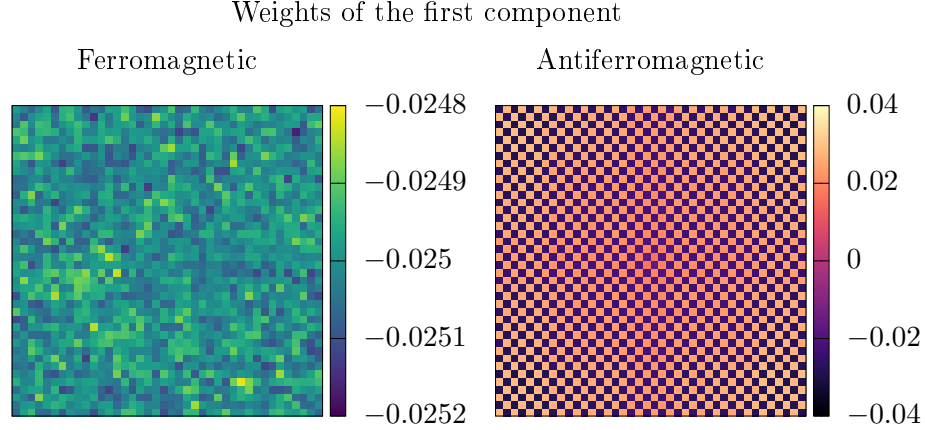


FIGURE 3 – Poids des premières composantes des deux systèmes de spin.

et phase ordonnée). On remarque dans la partie haute du graphe quelques échantillons à basse température et magnétisation nulle : il s'agit de ces états métastables où le système est coupé en deux régions uniformes d'orientation opposée.

Prédiction des phases On cherche maintenant à identifier les phases des configurations ferromagnétiques, projetées selon leur composante principale et re-normalisées (variance unité et moyenne nulle). On utilise pour cela une régression logistique. Une régression directe sur la première composante donne des résultats médiocres, mais (comme on s'y attend physiquement) la régression effectuée sur la valeur absolue de la magnétisation est très efficace (95% de réussite sur l'ensemble de test, avant optimisation). L'optimisation est réalisée par GridSearchCV sur le paramètre de régularisation C , sur une plage logarithmique de 10^{-2} à 10^2 . On compare ensuite les résultats optimaux selon le nombre de composantes principales retenues, entre 1 et 3 (Table 2). Les meilleurs résultats sont obtenus avec deux composantes : la magnétisation seule ne sait pas identifier les configurations ordonnées à deux régions, et la troisième composante conduit à de l'overfitting (moins bonne performance de généralisation à 3 composantes).

Résultats de la PCA sur le système antiferromagnétique En ce qui concerne les échantillons anti-ferromagnétique, l'analyse est moins efficace

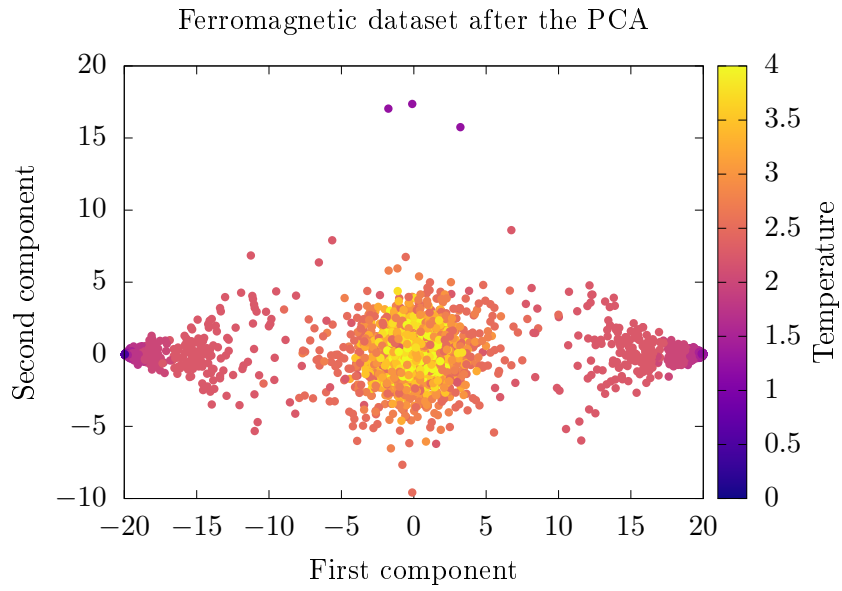


FIGURE 4 – Échantillons ferromagnétiques selon leurs deux composantes principales.

Number of components	1	2	3
C parameter	0.01	0.43	56.90
Accuracy – Validation set (%)	99.86	100.00	100.00
Accuracy – Test set (%)	95.46	96.26	91.24

TABLE 2 – Optimisation de la régression logistique (système ferromagnétique).

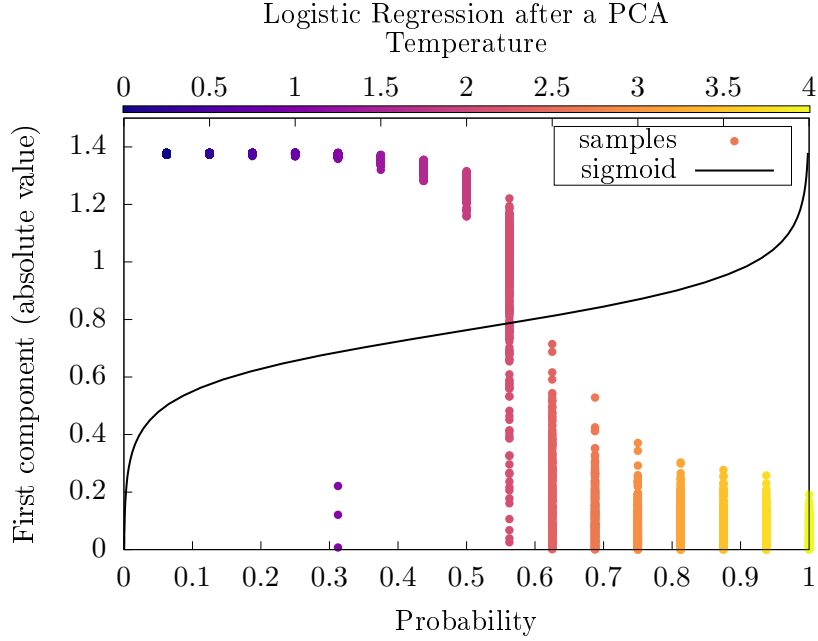


FIGURE 5 – La fonction logistique après optimisation sur les données ferromagnétiques.

Number of components	1	2	3
C parameter	0.72	2.68	11.79
Accuracy – Validation set (%)	89.50	91.67	96.79
Accuracy – Test set (%)	79.17	80.52	76.67

TABLE 3 – Optimisation de la régression logistique (système anti-ferromagnétique).

(Table 1). La première composante explique environ un tiers de la variance totale. Les poids de cette première composante, qui forment un damier, reflètent ici encore l'organisation des spins dans la phase ordonnée (Figure 3). La performance moindre de la PCA sur cet ensemble est illustrée Figure 6 : les configurations froides sont réparties sur les deux composantes ; seules les configurations chaudes sont concentrées en un amas central. Cela se retrouve sur les résultats de la régression logistique (3) : contrairement au cas ferromagnétique, l'ensemble de validation n'est pas parfaitement classé. La précision s'accroît avec le nombre de composantes considérées ; toutefois comme dans le cas précédent l'ensemble de test est le mieux réussi avec deux composantes principales. On l'interprète là encore comme le compromis entre information fournie et over-fitting de l'ensemble d'entraînement.

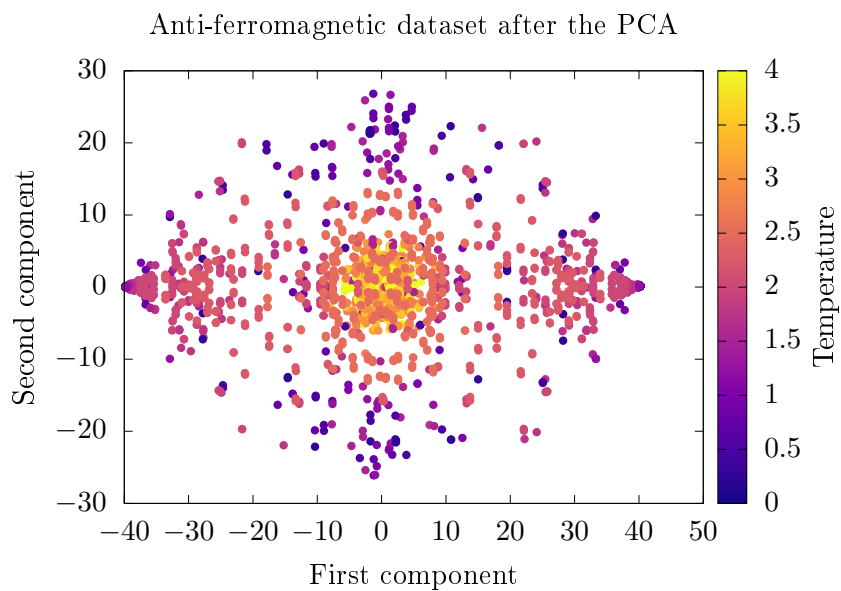


FIGURE 6 – Échantillons anti-ferromagnétiques selon leurs deux composantes principales.

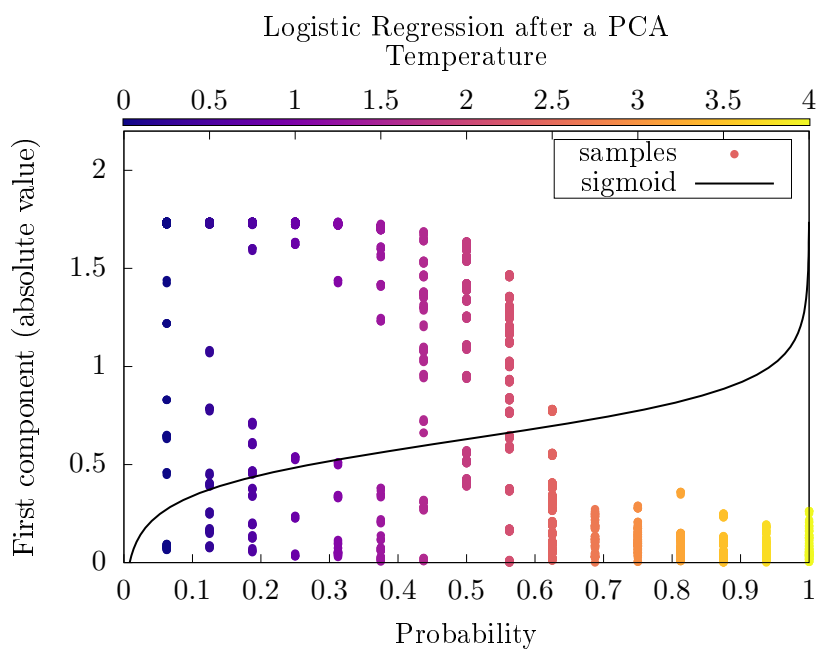


FIGURE 7 – La fonction logistique selon la première composante après optimisation sur les données antiferromagnétiques.

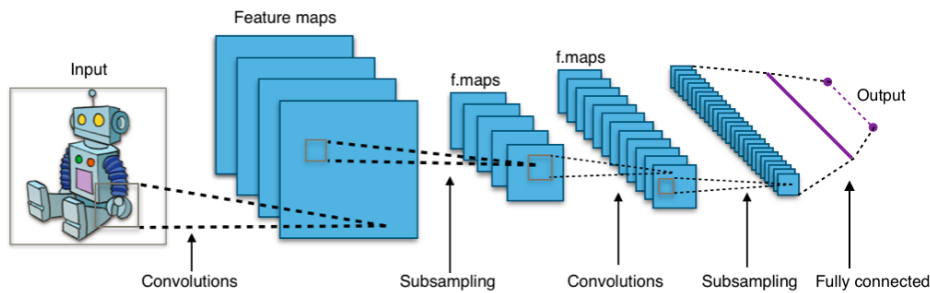


FIGURE 8 – Structure typique d'un réseau convolutif.[1]

4 Convolutional Neural Net (CNN) – Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs Il s'agit d'un réseau de neurones artificiels employé en analyse d'images. Ils sont formés d'une succession de couches de convolutions (Figure 8). Chaque couche est constituée de plusieurs noyaux de convolution, suivis d'un pooling qui réduit la dimension de l'image en sortie. Le réseau se termine habituellement par une couche de neurones complètement connectés donnant sur la couche de sortie.

Ces réseaux de neurones sont conçus pour prendre en compte l'organisation spatiale, locale, des images. Les noyaux de convolution des premières couches analysent des zones réduites de l'image. Leur résultat est agrégé par pooling. À chaque couche de convolution, la zone analysée correspond ainsi à une surface croissante de l'image d'entrée. Par rapport à un réseau multicouche de perceptrons complètement connectés, ils possèdent un nombre réduit de paramètres.

Le modèle Ici on choisit des noyaux de convolution carré car il n'y a a priori aucune raison de préférer un format différent. Leur nombre et leur taille sont des paramètres du modèle. Le pooling est composé de filtres max pooling de forme (2,2). Le modèle est complété par une couche complètement connectée de 32 neurones, et par un neurone final activé par une fonction sigmoïde (classification). Mis à part cette dernière couche, la fonction d'activation est une ReLU (Rectified Linear Unit) ; ce choix, de même que le max pooling, semble le plus répandu [8].

Concernant le nombre de couches, une règle générale est que la complexité des caractéristiques apprises par le modèle augmente avec ce nombre. Toutefois un modèle trop complexe mène à de l'overfitting de l'ensemble d'entraînement. Sur le problème qui nous occupe, une ou deux couches convolutives doivent suffire.

Le modèle est compilé avec la fonction de loss BinaryCrossentropy (adaptée à la classification binaire) et l'algorithme d'optimisation (optimiseur)

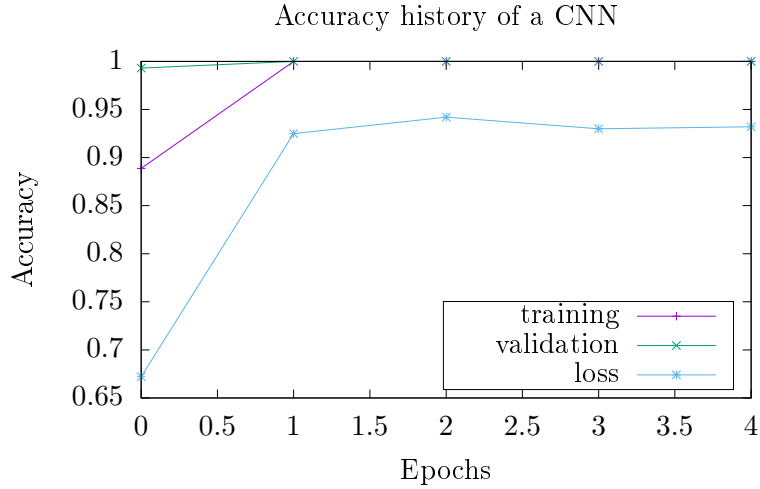


FIGURE 9 – Le réseau atteint très rapidement une grande précision sur les ensembles d’entraînement et de validation. En revanche il stagne sur l’ensemble de test après quelques epochs supplémentaires. L’architecture employée consiste en 1 couche de 6 noyaux de taille (4, 4).

Adam. Son taux d’apprentissage (learning rate) est un paramètre du modèle.

Optimisation Les échantillons sont remis sous leur forme spatiale 40×40 . On vérifie que leurs valeurs sont normalisées, c’est-à-dire qu’elles sont comprises entre 0 et 1. Le batch size est fixé à 64.

Quelques essais montrent que le CNN apprend très (trop) facilement à classer les échantillons de l’ensemble d’entraînement (100% de réussite Figure 9). Plutôt qu’optimiser les performances déjà maximales du modèle sur l’ensemble d’entraînement, on cherchera à quel point il est capable de généralisation sur l’ensemble de test. Pour cela on construit un callback Monitor() qui surveille les performances du modèle sur l’ensemble de test.

La difficulté principale réside dans la non-convexité de la réponse (précision) du modèle en fonction des hyper-paramètres. Les méthodes d’optimisation comprennent la recherche manuelle, parcourir une grille d’hyper-paramètres (GridSearch), des essais aléatoires (RandomizeSearch) ou encore des méthodes plus poussées d’inspiration bayésienne [6]. Une difficulté supplémentaire provient de la variabilité des performances, de l’ordre de 10%, pour des modèles construits à partir des mêmes hyper-paramètres (Figure 10). Il faut donc classer les modèles selon leur performance moyenne (CrossValidation). En pratique, on mesure la performance moyenne sur les ensembles de validation et de test de 5 modèles identiques. On étudie ainsi le learning rate, le nombre de couches (1 ou 2) et leur nombre de noyaux de

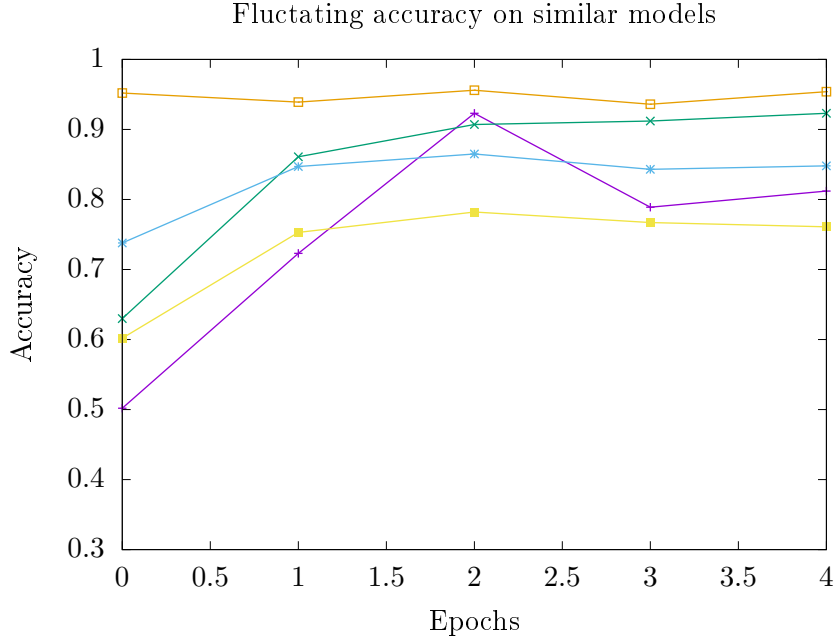


FIGURE 10 – Cinq modèles construits à partir des mêmes hyper-paramètres aboutissent à des performances diverses. Ici encore on observe une saturation du modèle. L’architecture employée consiste en 1 couche de 6 noyaux de taille (4, 4).

convolution, ainsi que la taille desdits noyaux.

Learning rate Le learning rate caractérise le pas de progression de l’algorithme Adam qui est utilisé pour ajuster les poids des neurones par descente de gradient. Un learning rate faible permet d’approcher les minima locaux du gradient, mais demande un grand nombre de pas. Au contraire un learning rate élevé progresse plus vite mais ne converge pas forcément. On utilise le callback `EarlyStopping()` pour ajuster la durée d’entraînement aux performances du modèle sur l’ensemble d’entraînement (on cherche le plateau à 100% de réussite). L’entraînement est ensuite prolongé pendant deux epochs pour améliorer les résultats sur l’ensemble de test (voir Figure 9).

Les meilleures performances sont obtenues pour la valeur de 10^{-4} . Elle présente aussi la durée moyenne d’apprentissage la plus courte. Bien qu’un résultat semblable ait été obtenu sur quelques architectures différentes (à 2 couches ou avec plus de noyaux), il n’est pas généralisable a priori. Il faudrait refaire cette analyse pour chaque modèle testé.

Architecture On commence par les modèles à une couche de convolution (Figure 5). La comparaison des tailles de noyaux révèle qu’un noyau (4, 4)

Learning rate	Validation set	Test set	Epochs
10^{-5}	100.00 (0.00)	83.38 (0.84)	9.0 (8.1)
10^{-4}	100.00 (0.00)	84.42 (1.07)	6.2 (3.4)
10^{-3}	100.00 (0.00)	82.30 (4.19)	12.2 (6.5)
10^{-2}	99.98 (0.04)	81.76 (7.45)	9.4 (5.9)

TABLE 4 – Précision (en %) sur les ensembles de validation et de test en fonction du learning rate. L'écart-type est indiqué entre parenthèse. La dernière colonne précise la durée moyenne d'entraînement.

Taille	Nombre de noyaux					
	1	2	4	6	8	16
2	71.4 (5.4)	76.1 (11.6)	83.00 (6.5)	80.6 (4.7)	84.0 (1.8)	86.4 (3.1)
4	71.5 (2.2)	80.0 (2.6)	81.3 (3.5)	82.5 (3.5)	85.2 (1.2)	86.9 (1.8)

TABLE 5 – Modèles à une couche. Les valeurs du tableau correspondent à la précision du CNN sur l'ensemble de test ; l'écart-type est indiqué entre parenthèses. On a par ailleurs 100% de précision sur l'ensemble de validation.

tend à donner de meilleurs résultats qu'un noyau (2,2). Au vu des écarts-types, ils semblent également plus stables. Enfin, sur ce modèle à une seule couche de convolution, la précision augmente avec le nombre de noyaux.

Concernant les modèles à deux couches, la première chose que l'on remarque, c'est la performance imparfaite des modèles sur l'ensemble de validation. Cela peut correspondre à quelques entraînements interrompus trop tôt par EarlyStopping(). Il est difficile par ailleurs de dégager des tendances de ce tableau. Le modèle [8, 16] est le plus performant des quatre.

Il ressort de ces quelques mesures que le réseau convolutif est difficile à optimiser sur les données inconnues du voisinage de la transition.

Taille	Architecture			
	[1, 8]	[4, 4]	[8, 1]	[8, 16]
2	90.4 (19.1)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
	72.3 (5.4)	80.1 (5.5)	72.5 (1.9)	84.9 (2.5)
4	100.0 (0.0)	95.0 (9.9)	85.5 (19.2)	100.00 (0.0)
	74.6 (4.0)	78.6 (7.1)	68.9 (3.0)	83.1 (3.1)

TABLE 6 – Modèles à deux couches. Le nombre de noyaux dans la première et dans la seconde couche sont indiquées entre crochets. Dans le tableau, La valeur supérieure correspond à la précision sur l'ensemble de validation ; celle située en dessous à l'ensemble de test. L'écart-type est précisé entre parenthèses.

5 Conclusion

Nous avons étudié deux types de modèles sur la classification des phases du modèle d'Ising.

Le premier repose sur l'analyse des composantes principales. Cette méthode permet de retrouver le paramètre d'ordre des transitions de phase para/ferro magnétique (magnétisation) et para/anti-ferro magnétique (magnétisation alternée). Une régression logistique sur les deux premières composantes donne des résultats impressionnants à proximité de la transition ferromagnétique. Ce travail peut être poursuivi dans la direction de l'apprentissage non-supervisé, à travers les méthodes de clustering.

Le second modèle étudié est un réseau de neurones convolutif qui vise à exploiter les corrélations spatiales entre spin. Il se révèle extrêmement performant sur les échantillons de phase déterminée, mais moins fiable sur les données inconnues près de la température critique. Il mériterait une optimisation plus complète (RandomSearch), qui ne fut pas implémentée par manque de puissance et temps.

Une approche supplémentaire consiste à construire des modèles génératifs : méthodes variationnelles, auto-encodeurs ou machines de Boltzmann. Au delà de la prédiction, ils permettraient éventuellement de comprendre certains aspects physiques (hamiltonien ou probabilités) du système étudié sous-jacent.

Références

- [1] Aphex34. Typical CNN. <https://commons.wikimedia.org/w/index.php?curid=45679374>, 2015. Consulté en jan 2021. Licence CC BY-SA 4.0.
- [2] A. Athenodorou and M. Sklodowska. Unsupervised recognition of phase transition on the lattice. https://indico.cern.ch/event/754973/contributions/3537862/attachments/1899967/3135914/Deep_Learning_Auto-encoder_Ising_Model.pdf, août 2019. Consulté en jan 2021.
- [3] W. Krauth. *The Ising model – Monte Carlo algorithms. Local sampling methods*, pages 249–252. Oxford University Press, New York, United States, 2006.
- [4] P. Mehta, M. Bukov, Ching-Hao Wang, A. G. R. Day, C. Richardson, C. Fisher, and D. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810 :1–124, 2019.
- [5] Pankaj Mehta. Ising MC. <https://physics.bu.edu/~pankajm/ML-Review-Datasets/isingMC/>, 2018. Consulté en oct 2020.

- [6] NeuPy. Hyperparameter optimization for neural networks. http://neupy.com/2016/12/17/hyperparameter_optimization_for_neural_networks.html, 2016. Consulté en janv 2021.
- [7] Wikipedia. Principal component analysis. https://en.wikipedia.org/wiki/Principal_component_analysis. Consulté en dec 2020.
- [8] Matthew D. Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks : 1st international conference on learning representations, iclr 2013. jan 2013. 1st International Conference on Learning Representations, ICLR 2013 ; Conference date : 02-05-2013 Through 04-05-2013.