



Tutorat Lyon Est

Unité d'Enseignement 3

BANQUE DE QCM

2020 - 2022

Intelligence Artificielle
QUESTIONS-REPONSES

QCM 1 – Cochez la (ou les) réponse(s) vraie(s) :

Concernant le perceptron monocouche :

- A. Il ne présente aucun biais.
- B. Il est composé d'un réseau de neurones.
- C. La fonction d'activation d'Heaviside est une fonction dite en marche d'escalier ne pouvant prendre que deux valeurs : 1 ou -1.
- D. Soit un perceptron monocouche comprenant un seul neurone en sortie auquel est associée une fonction d'activation d'Heaviside de seuil $\theta=5$ recevant le signal $\sum_{i=0}^I w_i X_i = 7$, la valeur renvoyée par la fonction d'activation d'Heaviside est 1.
- E. Le neurone de sortie reçoit directement le signal sous forme linéaire issu des neurones d'entrée.

QCM 1 – Cochez la (ou les) réponse(s) vraie(s) : BDE

A FAUX Le perceptron monocouche possède bel et bien un biais qui est noté b.

B VRAI Plus précisément, il s'agit d'un ensemble de neurones d'entrée qui peut être connecté à un ou plusieurs neurones de sorties.

C FAUX La fonction d'activation d'Heaviside est bien une fonction en marche et ne pouvant prendre que deux valeurs, mais ces deux valeurs sont 0 et 1.

D VRAI On peut soit directement comparer θ avec la valeur prise par la forme linéaire du signal produit par les neurones d'entrée, ou alors calculé la valeur z.

En calculant z, on obtient : $z = (\sum_{i=0}^I w_i X_i) - \theta = 7 - 5 = 2$. On remarque que $z \geq 0$, donc la valeur du signal renvoyé par la fonction d'activation d'Heaviside vaut 1.

E VRAI

QCM 2 – Cochez la (ou les) réponse(s) vraie(s) :

Concernant la Big Data :

- A. Le domaine de la protéomique génère plus de données (en terme quantitatif) que le domaine de la génomique.
- B. Le biais d'optimisme est majoré par l'augmentation de la taille de l'étude d'identification.
- C. Les omics sont les données issues de la biologie moléculaire et de l'imagerie médicale et entrent dans la catégorie de grands nombres de variables.
- D. Le biais d'optimisme est la conséquence d'une sélection statistique.
- E. Une étude d'identification a pour but d'estimer d'identifier des biomarqueurs et d'estimer leur effet sur le diagnostic, le pronostic et la réponse thérapeutique.

QCM 2 – Cochez la (ou les) réponse(s) vraie(s) : CDE

A FAUX C'est l'inverse, le domaine de la protéomique génère quelques centaines de données simultanément, alors que le domaine de la génomique en génère plusieurs milliards. Cela paraît logique puisque les protéines sont issues de parties de gènes, eux-mêmes constituant une petite partie du génome.

B FAUX C'est encore une fois l'inverse. Pour minimiser le biais d'optimisme, on peut augmenter la taille de l'étude d'identification. On peut également utiliser des méthodes pénalisées.

C VRAI Les omics représentent ainsi les données de la génomique, de la transcriptomique, de la protéomique, et plus récemment de la radiomique

D VRAI Cf. Partie VIII- Big Data

E VRAI Cf. Partie VIII- Big Data

QCM 3 – Intelligence Artificielle :

- A. Un réseau neuronal multicouche comporte une ou plusieurs couches cachées.
- B. La fonction tangente hyperbolique est définie sur l'intervalle $[0 ; 1]$.
- C. L'ensemble image de la fonction sigmoïde (logistique) est l'intervalle $[0 ; 1]$.
- D. Les neurones de la couche J des réseaux récurrents peuvent transmettre des informations à la couche J-1.
- E. Dans le cas du perceptron monocouche, l'ajustement des poids se fait par des méthodes de répropagation du gradient.

QCM 3 – Intelligence Artificielle : ACDE

A VRAI Au contraire du perceptron monocouche ne comporte pas de couche cachée.

B FAUX La fonction tangente hyperbolique est définie sur $\mathbb{R}$ et son ensemble image est l'intervalle $]-1 ; 1[$.

C VRAI

D VRAI

E VRAI Le perceptron monocouche est un réseau à propagation avant, l'ajustement des poids se fait donc par des méthodes de rétropropagation du gradient.

QCM 4 – Intelligence Artificielle :

- A. La fonction ReLU est une fonction affine par morceaux.
- B. La fonction Softmax est une extension de la fonction logistique.
- C. Un réseau neuronal multicouche est toujours un réseau récurrent.
- D. Les valeurs initiales données aux poids sont aléatoires.
- E. Dans les méthodes de propagation du gradient on commence par corriger les poids de la dernière couche.

QCM 4 – Intelligence Artificielle : ABDE

A VRAI

B VRAI La fonction sigmoïde est aussi appelée la fonction logistique, la fonction Softmax est donc une extension de la fonction logistique/sigmoïde.

C FAUX Un réseau neuronal multicouche peut être à propagation avant ou récurrent.

D VRAI

E VRAI

QCM 5 – Modèle linéaire et méthode des moindres carrés ordinaires :

- A. Dans cette méthode des moindres carrés ordinaires nous cherchons à minimiser la somme des écarts e_i .
- B. Les estimations uniques des paramètres β_0 et β_1 fournies par la méthode des moindres carrés sont uniques.
- C. Ces estimations correspondent aux valeurs qui rendent la somme des carrés des écarts minimale. S
- D. Un estimateur représente la fonction qui associe les valeurs d'un paramètre à une population.
- E. Plus la dispersion des valeurs observées autour des valeurs prédites est forte, plus la part de la variabilité de la variable Y expliquée par X est forte.

QCM 5 – Modèle linéaire et méthode des moindres carrés ordinaires : BC

A FAUX Attention le professeur a bien insisté, nous cherchons à minimiser la somme des **carrés** des écarts.

B VRAI Les valeurs estimées β_0 et β_1 sont obtenues en annulant les dérivées partielles de la somme des carrés des écarts par rapport à ces deux paramètres, nous avons donc bien une solution unique pour ces deux valeurs.

C VRAI

D FAUX Un estimateur représente la fonction associant les valeurs d'un paramètre à un échantillon.

E FAUX Plus la dispersion des valeurs observées autour des valeurs prédites est forte, plus la part de la variabilité de la variable Y expliquée par X est faible et inversement.

QCM 6 – La vraisemblance mais pas que :

- A. La vraisemblance de la valeur d'un paramètre est la probabilité des données si le paramètre a cette valeur.
- B. Dans le cas d'un modèle non linéaire, la méthode du maximum de vraisemblance donne les mêmes estimations des paramètres que la méthode des moindres carrés.
- C. La vraisemblance et la méthode des moindres carrés ordinaires sont deux méthodes pour estimer les paramètres d'un modèle.
- D. La vraisemblance est une méthode générale de construction d'estimateurs.
- E. Les résidus e_i correspondent aux écarts entre les valeurs observées y_i et les valeurs prédites par le modèle μ_i .

QCM 6 – La vraisemblance mais pas que : ACDE

A VRAI

B FAUX Attention, dans le cas d'un modèle non linéaire vous ne pouvez pas utiliser la méthode des moindres carrés ordinaires, seulement la méthode du maximum de vraisemblance. Pour un modèle linéaire, vous obtiendriez bien les mêmes estimations des paramètres.

C VRAI

D VRAI

E VRAI

OCM 7 – Intelligence Artificielle :

- A. Le rappel, aussi appelé spécificité, représente le nombre de vrais positifs sur le nombre total de malades.
- B. Les critères de performance d'un test diagnostique sont évalués à partir de l'échantillon d'apprentissage.
- C. La propriété de généralisation correspond à la capacité du système à classifier correctement des éléments qu'il découvre pour la première fois.
- D. Le deep learning concerne des réseaux de neurones monocouches.
- E. De nos jours, il est impossible pour une intelligence artificielle de surpasser l'être humain dans le domaine des jeux.

OCM 7 – Intelligence Artificielle : C

A FAUX La phrase est correcte à part pour le terme de spécificité. Le rappel correspond à la sensibilité.

B FAUX Attention, ces critères sont justement déterminés à partir de l'échantillon TEST ; non pas celui sur lequel l'IA s'est entraînée (risque de biais d'optimisme).

C VRAI

D FAUX Cela ne concerne que les réseaux multicouches.

E FAUX Cela est possible, notamment grâce à l'apprentissage par renforcement.

Question 8 – brr brr brr j'en peux plus de l'IA :

Soit un perceptron monocouche avec deux neurones en entrée et un neurone en sortie.

- A. Il y a plus de poids associés que de paramètres d'intérêt.
- B. On associe souvent à ce type de réseau la fonction d'activation d'Heaviside.
- C. Il est aussi possible d'associer ce perceptron à une fonction d'activation continue.
- D. Un perceptron monocouche peut comporter des couches cachées.
- E. Le perceptron finit toujours, à force d'essais, par trouver les poids que l'on avait choisi nous-même.

Question 8 – brr brr brr j'en peux plus de l'IA : ABC

A VRAI Dans ce cas, nous avons 2 neurones en entrée soit 2 paramètres d'intérêt. Chacun de ces paramètres est forcément associé à un poids, ce qui fait déjà deux poids. Or, on a également un poids, qu'on appelle w_0 , qui correspond en réalité au biais du modèle.

B VRAI C'est la fonction d'activation la plus utilisée pour les perceptrons monocouches.

C VRAI C'est d'ailleurs ce que le professeur a utilisé dans sa simulation (exemple du cours).

D FAUX C'est justement le principe du perceptron MONOCouche : pas de couches cachées !

E FAUX Cela dépend de la situation mais par exemple, dans le cas où le perceptron doit trouver une droite, identique à celle que nous aurions trouvée (c'est-à-dire qu'il n'a qu'une seule possibilité de réussite), il peut quand même trouver des poids différents de ceux auxquels nous avons pensé. En effet, les poids servent à calculer les paramètres de l'équation de la droite, or, ils sont utilisés dans des

rapports, ainsi, il suffit que le rapport final soit identique à celui que nous avons pour que la droite soit identique. Pour avoir un rapport identique, il est possible d'avoir des poids différents (par exemple, 2/3 et 4/6 donneront les mêmes rapports avec pourtant des valeurs différentes !).

Question 9 – allez encore un petit effort :

- A. Dans le modèle linéaire, Y explique toute la variabilité de X.
- B. La méthode des moindres carrés ordinaires ne s'applique que pour les modèles linéaires.
- C. La méthode du maximum de vraisemblance ne s'applique que dans les modèles linéaires.
- D. Un réseau neuronal peut prédire toute la variabilité d'un paramètre.
- E. L'apprentissage profond se fait par modification progressive du poids des paramètres.

Question 9 – allez encore un petit effort : BE

A FAUX Une partie de la variabilité de X n'est pas expliquée par Y, c'est ce qu'on appelle la résiduelle.

B VRAI

C FAUX Cette méthode n'est pas spécifique d'un modèle : on peut l'appliquer pour tous les modèles (survie, logistique, linéaire).

D FAUX Un réseau neuronal ne fait pas mieux qu'un modèle, ainsi, il y a toujours une résiduelle, comme à l'item A. Un réseau neuronal est limité dans ses prédictions par cette part de variabilité expliquée par le modèle.

E VRAI

Question 10 – ok c'est mon dernier, le prochain c'est à l'EB :

- A. Un réseau neuronal multicouche corrige un nombre plus important de poids qu'un réseau monocouche.
- B. La fonction sigmoïde est intéressante pour la prédiction de probabilités.
- C. La discontinuité de la fonction d'Heaviside permet une meilleure correction des poids.
- D. L'exactitude d'un réseau se calcule en utilisant cette formule :

$$\frac{VN + FN}{VN + FN + VP + FP}$$

- E. Un biomarqueur est une variable comme la douleur, qui peut alors servir de biomarqueur de pronostic.

Question 10 – ok c'est mon dernier, le prochain c'est à l'EB : AB

A VRAI On peut donc réaliser des prédictions impliquant plus de variables.

B VRAI Elle est définie sur l'intervalle [0 ;1] comme les probabilités ce qui est pratique pour ces prédictions.

C FAUX Justement non ! Il est beaucoup plus facile de corriger des poids avec une fonction continue comme la fonction sigmoïde par exemple.

D FAUX Dans la formule de l'exactitude, il faut mettre au dénominateur toutes les valeurs « vraies » donc les vrais positifs et les vrais négatifs, et au numérateur, il faut mettre toute la population (vrais et faux négatifs, vrais et faux positifs).

$$Exactitude = \frac{VP + VN}{VP + FP + VN + FN}$$

E FAUX Un biomarqueur est une variable BIOLOGIQUE. La douleur est un signe fonctionnel et pas une variable biologique, cela ne peut donc pas constituer un biomarqueur 😊.

Question 11 – Santired :

Le Père Noël est épuisé, Roger est le seul lutin qui n'est pas tombé malade et le Père Noël est donc obligé de fabriquer les jouets à ses côtés pour être prêt le 24 décembre. Pour avoir un peu d'aide, il décide de mettre au point un perceptron monocouche, composé de 3 neurones d'entrée et d'un neurone de sortie, à propagation avant, qui lui permettra, à partir d'informations collectées sur l'enfant en question, de faire un choix entre un robot télécommandé et de la pâte à modeler. Le Père Noël vérifie ensuite si le choix est adapté et aide Roger à la construction : c'est un réel gain de temps, il ne réfléchit pas des heures sur le choix du cadeau adapté à l'enfant.

- A. Le neurone de sortie associe un poids à chacune des 3 informations qu'il reçoit.
- B. Sur le plan théorique, la fonction d'Heaviside est adaptée à cette situation puisqu'elle renvoie un signal Booléen.
- C. Si le Père Noël avait voulu un réseau plus compliqué, il aurait pu choisir un réseau neuronal multicouche, comprenant moins de neurones en entrée que de signaux captés.
- D. Si le Père Noël construit d'autres réseaux, il pourra les entraîner grâce à la fonction d'activation SoftMax.
- E. Les algorithmes d'apprentissage par rétropropagation à travers le temps sont utilisés pour l'entraînement des réseaux récurrents.

Question 11 – Santired : ABDE

A VRAI Le neurone de sortie reçoit des informations de chacun des 3 neurones d'entrée : ces informations sont notées X_i (i prenant les valeurs 1 à 3). A chacune d'entre elles, il associe un poids w_i , conduisant à la somme pondérée suivante (forme linéaire) :

$$\sum_{i=0}^I w_i X_i$$

(Dans cette somme, w_0 correspond à la valeur du biais).

B VRAI Ici, le perceptron doit permettre de faire un choix entre 2 options : la fonction d'Heaviside est parfaitement adaptée sur le plan théorique puisqu'elle donne uniquement 2 valeurs en sortie (c'est ce que signifie un « signal Booléen ») selon que la valeur de la forme linéaire soit supérieure ou inférieure au seuil choisi. Sur le plan pratique, d'autres fonctions sont utilisées pour pallier les contraintes liées à la discontinuité de la fonction Heaviside (voir cours de cette année).

C FAUX Le principe d'un réseau neuronal est d'avoir autant de neurones en entrée que de signaux à capter.

D VRAI La fonction Softmax est une extension de la fonction sigmoïde. C'est une fonction continuellement dérivable qui peut donc être utilisée pour l'entraînement des réseaux.

E VRAI Cf cours.

Pour les réseaux neuronaux à propagation avant, les corrections des poids se font par des méthodes de rétropropagation du gradient ou par d'autres méthodes assimilées. On corrige d'abord les poids de la dernière couche, puis ceux des couches intermédiaires, pour finir par corriger ceux de la première couche. Pour les réseaux récurrents, les corrections des poids se font par des méthodes de rétropropagation à travers le temps.

Question 12 – Lucien Lutin n'y comprend rien :

Lucien Lutin est de nouveau sur pied, il revient donc aider Roger et le Père Noël à préparer les cadeaux. Mais face à l'Intelligence Artificielle (IA) que le Père Noël a programmée, il ne sait pas où donner de la tête. Il émet alors un certain nombre d'affirmations dont on vous demande de vérifier la véracité.

- A. Grâce à l'apprentissage supervisé, le réseau neuronal apprend de ses erreurs.
- B. Dans l'apprentissage profond, on entraîne l'IA sur une centaine d'exemples.
- C. Le Big Data a toute sa place en médecine dans le domaine génomique.
- D. Les modèles diagnostiques étudient l'influence des covariables sur la réponse au traitement.
- E. Pour estimer les paramètres d'un modèle linéaire, on pourra utiliser la méthode des moindres carrés ordinaires.

Question 12 – Lucien Lutin n'y comprend rien : ACE

A VRAI Dans l'apprentissage supervisé, le réseau de neurone apprend à prédire la valeur la plus probable d'une variable Y en fonction des valeurs des covariables X. Pour les réseaux neuronaux, l'apprentissage supervisé se fait en les entraînant à reconnaître des images, des objets, des visages... L'algorithme apprend de ses erreurs. A chaque erreur, les paramètres (les poids) sont progressivement ajustés. Ainsi, lorsque l'algorithme se trouvera confronté à nouveau à cet objet, il y aura une forte probabilité qu'il le reconnaisse.

B FAUX Il faut bien comprendre que l'IA nécessite des dizaines de milliers d'exemple pour pouvoir « apprendre » : une centaine d'exemples ne pourrait pas suffire à une IA donnant des résultats convenables.

C VRAI Oui complètement ! De même pour les domaines transcriptomiques et protéomique.

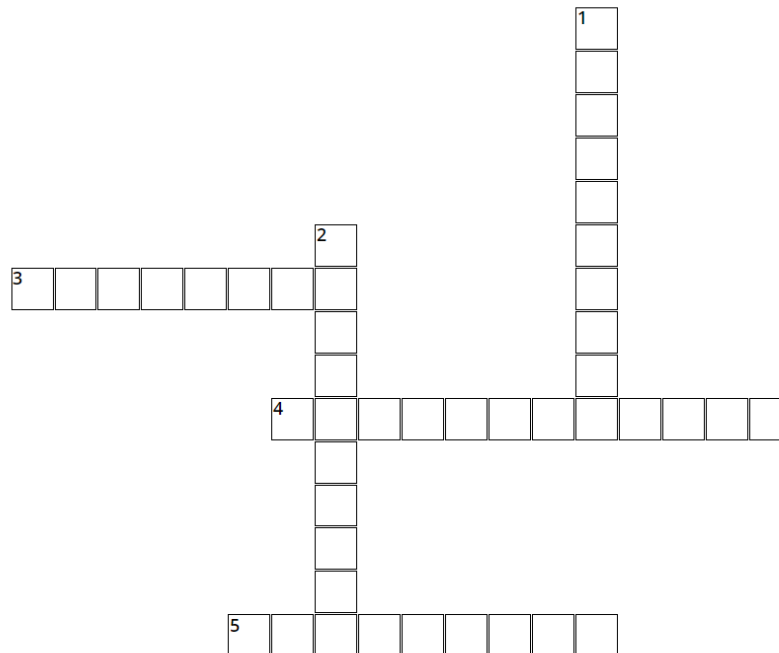
D FAUX Non ! C'est la définition des modèles théranostiques.

E VRAI L'objectif de cette méthode est de minimiser la somme des écarts quadratiques (au carré) entre les valeurs observées y_i et les valeurs prédites μ_i sur l'ensemble des valeurs de l'échantillon. Dans le cas d'un modèle linéaire simple univarié, cette méthode fournit des estimations de l'ordonnée à l'origine et de la pente de la droite de régression.

Question 13 – Petite pause mots croisés sur la plage :

Nos chers P1 ont également acheté un magazine de mots croisés au tabac du coin de la plage, aidez-les à remplir les cases :

INTELLIGENCE ARTIFICIELLE



Horizontal

3. Fonction d'activation définie sur l'intervalle $[0;1]$
4. Méthode d'apprentissage utilisée pour les jeux, dont les performances dépassent celles humaines: apprentissage par "..."
5. La méthode des moindres carrés ordinaires n'est utilisable que dans un cas de modèles. Ce sont les modèles dits "..."

- A. 1 = régressive.
- B. 2 = Heavisides.
- C. 3 = tangente.
- D. 4 = renforcement.
- E. 5 = régressif.

Vertical

1. La variance non expliquée par le modèle est appelée variance "..."
2. Ensemble de neurones sans couche cachée (en deux mots, je ne vous demande que le premier)

Question 13 – Petite pause mots croisés sur la plage : D

A FAUX Un modèle n'explique jamais complètement la variabilité d'un paramètre (tout comme le temps n'expliquait pas toute la variabilité de la quantité de sable dans l'exo d'avant). La variance non expliquée est appelée variance résiduelle.

B FAUX On nous donne ici la définition du perceptron monocouche (notion à bien revoir car très importante pour le professeur).

C FAUX Explications détaillées. C'est la fonction sigmoïde. En effet, la fonction tangente est définie sur $[-1;1]$. La fonction sigmoïde est très pratique pour modéliser des probabilités.

D VRAI Exactement. Il en existe d'autres : le deep learning, ou l'apprentissage supervisé. Toutes ces méthodes demandent énormément d'entraînements, donc cela nécessite d'avoir accès à des montagnes de données.

E FAUX Et non (en vrai, si vous regardiez bien, il n'y a pas de S à régressif alors que la phrase demande à être remplie au pluriel donc la réponse ne pouvait pas être vraie). Les moindres carrés ordinaires permettent de modéliser les estimateurs dans les modèles linéaires uniquement. Sinon, on pourra utiliser la méthode du maximum de vraisemblance.