Devoir 1

IFT714 : Traitement automatique des langues naturelles

Enseignant : Amine Trabelsi

Étudiant : Mamadou Senghor

CIP : senm1912

Université de Sherbrooke

Département d'informatique

Hiver 2026

Date de remise : 19 janvier 2026

Table des matières

1 Question 1 – Vrai ou Faux

Dans cette question, nous devons déterminer si chaque affirmation est vraie ou fausse. Si elle est vraie, nous fournissons une démonstration. Si elle est fausse, nous donnons un contre-exemple ou une explication détaillée, en incluant toutes les étapes intermédiaires de calcul, conformément aux consignes du devoir.

1.1 Affirmation (a)

Énoncé : Soit $A \in \mathbb{R}^{n\times n}$. On définit $X = A^T A$. L'affirmation est que X est une matrice symétrique.

Rappel : Une matrice X est symétrique si et seulement si $X^T = X$.

Preuve :
Calculons la transposée de X :

$$X^T = (A^T A)^T$$

En utilisant la propriété $(BC)^T = C^T B^T$, nous obtenons :

$$X^T = A^T (A^T)^T$$

Or, on sait que $(A^T)^T = A$. Donc :

$$X^T = A^T A$$

Ainsi :

$$X^T = X$$

Conclusion : L'affirmation est **vraie**. La matrice $X = A^T A$ est toujours symétrique. C'est une propriété générale : pour toute matrice A (pas nécessairement carrée), le produit $A^T A$ est toujours symétrique.

1.2 Affirmation (b)

Énoncé : Soit $X \in \mathbb{R}^{n\times n}$ et $\alpha \in \mathbb{R}^n$. On affirme que

$$\alpha^T X^T X \alpha \geq 0.$$

Preuve :
Posons :

$$v = X\alpha.$$

Alors, on peut réécrire l'expression donnée :

$$\alpha^T X^T X \alpha = (X\alpha)^T (X\alpha) = v^T v.$$

Or, par définition :

$$v^T v = \sum_{i=1}^{n} v_i^2.$$

Comme chaque terme $v_i^2 \geq 0$, on a :

$$v^T v \geq 0.$$

Donc :

$$\alpha^T X^T X \alpha \geq 0.$$

Conclusion : L'affirmation est **vraie**.

1.3 Affirmation (d)

Énoncé : Soit

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

On affirme que

$$A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

Rappel : Si A est diagonale, $A = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ avec $\lambda_i \neq 0$, alors

$$A^{-1} = \mathrm{diag}\Big(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \frac{1}{\lambda_3}\Big).$$

Application : Ici,

$$A = \mathrm{diag}(2, -2, 3),$$

donc

$$A^{-1} = \mathrm{diag}\Big(\frac{1}{2}, -\frac{1}{2}, \frac{1}{3}\Big) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

Vérification : On vérifie que $AA^{-1} = I$:

$$A\,A^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 2 \cdot \frac{1}{2} & 0 & 0 \\ 0 & (-2) \cdot \big(-\frac{1}{2}\big) & 0 \\ 0 & 0 & 3 \cdot \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I.$$

Conclusion : L'affirmation est **fausse**. En effet, le signe de l'élément diagonal central doit être négatif : la bonne réponse est

$$A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

2 Question 2 – Entropie

On rappelle que l'entropie d'une variable aléatoire discrète X est définie par :

$$H(X) = -\sum_x p(x) \log p(x),$$

où $p(x) = \mathbb{P}(X = x)$.

2.1 (a) Cas uniforme sur $\{1, 2, \ldots, m\}$

Si X est uniforme sur $\{1, 2, \ldots, m\}$, alors pour tout $x \in \{1, \ldots, m\}$:

$$p(x) = \frac{1}{m}.$$

Ainsi,

$$H(X) = -\sum_{x=1}^{m} \frac{1}{m} \log\left(\frac{1}{m}\right).$$

Comme le terme $\log\left(\frac{1}{m}\right)$ ne dépend pas de x, on le factorise :

$$H(X) = -\left(\sum_{x=1}^{m} \frac{1}{m}\right) \log\left(\frac{1}{m}\right).$$

Or,

$$\sum_{x=1}^{m} \frac{1}{m} = m \cdot \frac{1}{m} = 1.$$

Donc

$$H(X) = -\log\left(\frac{1}{m}\right) = \log(m).$$

Conclusion : $H(X) = \log(m)$.

2.2 (b) Cas Bernoulli(ϕ)

Si $X \sim$ Bernoulli(ϕ), alors

$$p(1) = \phi, \qquad p(0) = 1 - \phi.$$

Donc, par définition :

$$H(X) = -\Big(p(0) \log p(0) + p(1) \log p(1)\Big).$$

En remplaçant :

$$H(X) = -\Big((1 - \phi) \log(1 - \phi) + \phi \log(\phi)\Big).$$

Conclusion :

$$H(X) = -\left[\phi \log \phi + (1 - \phi) \log(1 - \phi)\right].$$

2.3 (c) Valeur de ϕ qui maximise $H(X)$

On considère

$$H(\phi) = -\left[\phi \log \phi + (1-\phi)\log(1-\phi)\right], \quad \phi \in (0,1).$$

Étape 1 : dérivée

On dérive terme à terme.
D'abord,

$$\frac{d}{d\phi}\Big(\phi \log \phi\Big) = \log \phi + 1.$$

Ensuite, pour $(1-\phi)\log(1-\phi)$, on utilise la règle du produit :

$$\frac{d}{d\phi}\Big((1-\phi)\log(1-\phi)\Big) = \frac{d}{d\phi}(1-\phi)\cdot\log(1-\phi) + (1-\phi)\cdot\frac{d}{d\phi}\log(1-\phi).$$

Or,

$$\frac{d}{d\phi}(1-\phi) = -1 \quad \text{et} \quad \frac{d}{d\phi}\log(1-\phi) = \frac{-1}{1-\phi}.$$

Donc,

$$\frac{d}{d\phi}\Big((1-\phi)\log(1-\phi)\Big) = (-1)\log(1-\phi) + (1-\phi)\left(\frac{-1}{1-\phi}\right) = -\log(1-\phi) - 1.$$

Ainsi,

$$H'(\phi) = -\left[(\log\phi + 1) + (-\log(1-\phi) - 1)\right] = -\left[\log\phi - \log(1-\phi)\right].$$

Donc

$$H'(\phi) = -\log\left(\frac{\phi}{1-\phi}\right).$$

Étape 2 : point critique

On impose $H'(\phi) = 0$:

$$-\log\left(\frac{\phi}{1-\phi}\right) = 0 \quad \Longleftrightarrow \quad \log\left(\frac{\phi}{1-\phi}\right) = 0 \quad \Longleftrightarrow \quad \frac{\phi}{1-\phi} = 1.$$

Donc

$$\phi = 1-\phi \quad \Longleftrightarrow \quad 2\phi = 1 \quad \Longleftrightarrow \quad \phi^{=\frac{1}{2}}.$$

Conclusion : $H(X)$ est maximisée pour $\phi^{=\frac{1}{2}}$.

2.4 (d) Montrer que si X et Y sont indépendantes, alors $H(X,Y) = H(X) + H(Y)$

On rappelle l'entropie jointe :

$$H(X,Y) = -\sum_x \sum_y p(x,y) \log p(x,y).$$

Si X et Y sont indépendantes, alors

$$p(x,y) = p(x)p(y).$$

En remplaçant dans la définition :

$$H(X,Y) = -\sum_x \sum_y p(x)p(y) \log \Big(p(x)p(y)\Big).$$

En utilisant $\log(ab) = \log a + \log b$:

$$H(X,Y) = -\sum_x \sum_y p(x)p(y) \Big(\log p(x) + \log p(y)\Big).$$

On sépare en deux sommes :

$$H(X,Y) = -\sum_x \sum_y p(x)p(y) \log p(x) \; - \; \sum_x \sum_y p(x)p(y) \log p(y).$$

Premier terme

$$-\sum_x \sum_y p(x)p(y) \log p(x) = -\sum_x \Big(p(x) \log p(x) \sum_y p(y)\Big).$$

Or, $\sum_y p(y) = 1$, donc

$$-\sum_x \sum_y p(x)p(y) \log p(x) = -\sum_x p(x) \log p(x) = H(X).$$

Deuxième terme

$$-\sum_x \sum_y p(x)p(y) \log p(y) = -\sum_y \Big(p(y) \log p(y) \sum_x p(x)\Big).$$

Or, $\sum_x p(x) = 1$, donc

$$-\sum_x \sum_y p(x)p(y) \log p(y) = -\sum_y p(y) \log p(y) = H(Y).$$

Conclusion finale

En combinant les deux résultats :

$$H(X,Y) = H(X) + H(Y).$$

3 Question 3 – Bernoulli : moyenne/variance et estimateur du maximum de vraisemblance

On suppose que $Y \sim \text{Bernoulli}(\phi)$, c'est-à-dire

$$\mathbb{P}(Y = 1) = \phi, \qquad \mathbb{P}(Y = 0) = 1 - \phi,$$

et que $D = \left(y^{(1)}, \ldots, y^{(m)}\right)$ est un échantillon i.i.d. tiré de Y.

3.1 (a) Calcul de $\mathbb{E}(Y)$ et $\text{Var}(Y)$

Espérance

Par définition,

$$\mathbb{E}(Y) = \sum_{y \in \{0,1\}} y\, \mathbb{P}(Y = y) = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1).$$

En remplaçant $\mathbb{P}(Y = 1) = \phi$:

$$\mathbb{E}(Y) = 0 \cdot (1 - \phi) + 1 \cdot \phi = \phi.$$

Conclusion : $\mathbb{E}(Y) = \phi$.

Variance

On utilise :

$$\text{Var}(Y) = \mathbb{E}(Y^2) - \left(\mathbb{E}(Y)\right)^2.$$

Comme $Y \in \{0, 1\}$, on a $Y^2 = Y$ (car $0^2 = 0$ et $1^2 = 1$). Donc :

$$\mathbb{E}(Y^2) = \mathbb{E}(Y) = \phi.$$

Ainsi,

$$\text{Var}(Y) = \phi - \phi^2 = \phi(1 - \phi).$$

Conclusion : $\text{Var}(Y) = \phi(1 - \phi)$.

3.2 (b) Estimateur du maximum de vraisemblance de ϕ

On veut maximiser la log-vraisemblance

$$\ell(\phi) = \log\left(\prod_{i=1}^{m} p\left(y^{(i)}\right)\right).$$

Étape 1 : écrire la vraisemblance

Pour une observation $y^{(i)} \in \{0, 1\}$, la loi Bernoulli s'écrit :

$$p\left(y^{(i)}\right) = \phi^{y^{(i)}}(1 - \phi)^{1 - y^{(i)}}.$$

Donc la vraisemblance est :

$$L(\phi) = \prod_{i=1}^{m} \phi^{y^{(i)}}(1 - \phi)^{1 - y^{(i)}}.$$

Étape 2 : passer au log

$$\ell(\phi) = \log L(\phi) = \log\left(\prod_{i=1}^{m} \phi^{y^{(i)}}(1-\phi)^{1-y^{(i)}}\right).$$

En utilisant $\log\left(\prod_i a_i\right) = \sum_i \log(a_i)$:

$$\ell(\phi) = \sum_{i=1}^{m} \log\left(\phi^{y^{(i)}}(1-\phi)^{1-y^{(i)}}\right).$$

En utilisant $\log(ab) = \log a + \log b$ et $\log(u^k) = k \log u$:

$$\ell(\phi) = \sum_{i=1}^{m} \left(y^{(i)} \log \phi + (1-y^{(i)}) \log(1-\phi)\right).$$

Étape 3 : dériver et annuler

On dérive :

$$\frac{d}{d\phi}\left(y^{(i)} \log \phi\right) = y^{(i)} \frac{1}{\phi}, \qquad \frac{d}{d\phi}\left((1-y^{(i)}) \log(1-\phi)\right) = (1-y^{(i)})\left(\frac{-1}{1-\phi}\right).$$

Donc :

$$\ell'(\phi) = \sum_{i=1}^{m} \left(\frac{y^{(i)}}{\phi} - \frac{1-y^{(i)}}{1-\phi}\right).$$

On pose $\ell'(\phi) = 0$.

Posons :

$$S = \sum_{i=1}^{m} y^{(i)}.$$

Alors :

$$\sum_{i=1}^{m} (1-y^{(i)}) = m - S.$$

Donc l'équation $\ell'(\phi) = 0$ devient :

$$\frac{S}{\phi} - \frac{m-S}{1-\phi} = 0 \quad \Longleftrightarrow \quad \frac{S}{\phi} = \frac{m-S}{1-\phi}.$$

On croise :

$$S(1-\phi) = \phi(m-S).$$

On développe :

$$S - S\phi = m\phi - S\phi.$$

On simplifie $-S\phi$ des deux côtés :

$$S = m\phi.$$

Ainsi,

$$\hat{\phi}_{\text{MLE}} = \frac{S}{m} = \frac{1}{m}\sum_{i=1}^{m} y^{(i)}.$$

Conclusion : L'estimateur du maximum de vraisemblance est

$$\boxed{\hat{\phi}_{\text{MLE}} = \frac{1}{m}\sum_{i=1}^{m} y^{(i)}.}$$

C'est simplement la moyenne empirique des observations.

4 Question 4 – Dérivées et gradients

4.1 (a) Dérivée de la sigmoïde

Énoncé : Soit

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

Montrer que

$$\frac{d}{dx}\sigma(x) = \sigma(x)\Big(1-\sigma(x)\Big).$$

Preuve (avec étapes) :
On réécrit la sigmoïde sous forme de puissance :

$$\sigma(x) = (1+e^{-x})^{-1}.$$

On dérive en utilisant la règle de dérivation de $u(x)^{-1}$:

$$\frac{d}{dx}\left(u^{-1}\right) = -u^{-2}\,u'.$$

Ici, $u(x) = 1+e^{-x}$, donc

$$u'(x) = \frac{d}{dx}(1+e^{-x}) = 0 + \frac{d}{dx}(e^{-x}) = -e^{-x}.$$

Ainsi,

$$\sigma'(x) = -(1+e^{-x})^{-2}\cdot(-e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^2}.$$

Maintenant, calculons $\sigma(x)(1-\sigma(x))$:

$$1-\sigma(x) = 1 - \frac{1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}}.$$

Donc,

$$\sigma(x)(1-\sigma(x)) = \frac{1}{1+e^{-x}}\cdot\frac{e^{-x}}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2}.$$

On obtient bien :

$$\sigma'(x) = \sigma(x)\Big(1-\sigma(x)\Big).$$

Conclusion : La relation est démontrée.

4.2 (b) Gradient de $g(x) = x^T w$

Énoncé : Soient $x = (x_1, x_2)^T \in \mathbb{R}^2$ et $w = (w_1, w_2)^T \in \mathbb{R}^2$. On définit

$$g(x) = x^T w.$$

Montrer que $\nabla_x g(x) = w$.

Calcul :

On développe :

$$g(x) = x^T w = x_1 w_1 + x_2 w_2.$$

Donc,

$$\frac{\partial g}{\partial x_1} = w_1, \qquad \frac{\partial g}{\partial x_2} = w_2.$$

Par définition,

$$\nabla_x g(x) = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = w.$$

Conclusion : $\nabla_x g(x) = w$.

4.3 (c) Gradient de $Q(x) = x^T Ax$ et comparaison avec $(A + A^T)x$

Énoncé : Soient $x = (x_1, x_2)^T \in \mathbb{R}^2$,

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \qquad Q(x) = x^T Ax.$$

Calculer $\nabla_x Q(x)$ et comparer à $(A + A^T)x$.

Étape 1 : calculer Ax

$$Ax = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ 2x_1 + x_2 \end{pmatrix}.$$

Étape 2 : calculer $Q(x) = x^T(Ax)$

$$Q(x) = (x_1, x_2) \begin{pmatrix} x_1 + x_2 \\ 2x_1 + x_2 \end{pmatrix} = x_1(x_1 + x_2) + x_2(2x_1 + x_2).$$

On développe :

$$Q(x) = x_1^2 + x_1 x_2 + 2x_1 x_2 + x_2^2 = x_1^2 + 3x_1 x_2 + x_2^2.$$

Étape 3 : gradient

$$\frac{\partial Q}{\partial x_1} = 2x_1 + 3x_2, \qquad \frac{\partial Q}{\partial x_2} = 3x_1 + 2x_2.$$

Donc :

$$\nabla_x Q(x) = \begin{pmatrix} 2x_1 + 3x_2 \\ 3x_1 + 2x_2 \end{pmatrix}.$$

Étape 4 : comparer avec $(A + A^T)x$

On calcule :

$$A^T = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \quad \Rightarrow \quad A + A^T = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}.$$

Alors :

$$(A + A^T)x = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + 3x_2 \\ 3x_1 + 2x_2 \end{pmatrix}.$$

On observe donc :

$$\nabla_x Q(x) = (A + A^T)x.$$

Conclusion : $\nabla_x (x^T Ax) = (A + A^T)x$ (ici, les deux coïncident exactement).

4.4 (d) Déduire $\nabla_x \|x\|_2^2$

Énoncé : En utilisant le résultat de (c), déduire

$$\nabla_x \|x\|_2^2, \qquad \text{où } \|x\|_2^2 = x^T x.$$

Étape 1 : écrire $x^T x$ sous la forme $x^T A x$
On remarque que

$$x^T x = x^T I x,$$

où I est la matrice identité 2×2 (ou $n \times n$ en général).

Étape 2 : appliquer le résultat de (c)
D'après (c), pour $Q(x) = x^T A x$,

$$\nabla_x Q(x) = (A + A^T)x.$$

Ici, $A = I$ et $I^T = I$, donc :

$$(A + A^T)x = (I + I)x = 2Ix = 2x.$$

Conclusion :

$$\boxed{\nabla_x \|x\|_2^2 = 2x.}$$