IFT714: Traitement automatique des langues naturelles

Dû, le 19 janvier 2026
23h59 heure de l'est

Devoir #1

Hiver 2026

Enseignant: Amine Trabelsi

- Ce devoir comprend quatre (4) problèmes. C'est un bref rappel de certains résultats de l'algèbre matricielle, de la statistique et du calcul différentiel. Les réponses aux questions devraient inclure les étapes de calcul intermédiaires qui vous ont mené à la réponse suggérée. Aucun crédit ne sera accordé aux réponses qui ne répondent pas à ce critère.
- L'utilisation de LaTeX est préférable mais pas obligatoire. Vous pouvez consulter un tutoriel LaTeX et ce cheatsheet. Dans la plupart des cas, si vous voulez écrire quelque chose en LaTeX, vous pouvez simplement Google "how to do {X} in latex" et les premiers liens devraient fournir la syntaxe que vous recherchez.
- Incluez votre nom et CIP avec votre soumission.
- Toutes les soumissions doivent être au format PDF suivant ce format: "prenom_nom_Devoir1_CIP".
- Si vous souhaitez soumettre des réponses manuscrites, vous pouvez les numériser et les soumettre au format PDF.
- Le devoir doit être remis sur le site du cours sur Moodle avant 23h59 heure de l'est à la date d'échéance.
- Selon la politique de jours de retard énoncée sur le plan de cours, un devoir soumis 24 heures après la date limite sera pénalisé de 3%. Un devoir soumis deux jours (24 à 48 heures) après la date limite sera pénalisé de 10%, et un devoir soumis trois jours sera pénalisé de 20%. Soumettre une livrable 72 heures (3 jours) après la date limite ne sera pas accepté.
- Les soumissions incomplètes ou en retard seront évaluées en tant que telles et aucun accommodement ou réévaluation ne sera fait.

---

1. Indiquez laquelle des affirmations suivantes est vraie ou fausse. Si vrai, expliquez pourquoi et si faux, fournissez un contre-exemple ou expliquez. (1 + 2 + 1 + 1 = 5 Points)
   (a) $A \in \mathbb{R}^{n\times n}, X = A^T A$ est une matrice symétrique.
   (b) $X \in \mathbb{R}^{n\times n}, \boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\alpha}^T X^T X \boldsymbol{\alpha} \geqslant 0$
   (c) $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

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

2. L'entropie d'une variable aléatoire discrète X est définie comme : (1 + 1.5 + 1.5 + 2 = 6 Points)

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

où $p(x) = P(X = x)$, et $H(X)$ est interprété comme l'information moyenne révélée par un résultat aléatoire.

(a) Calculez l'entropie d'une variable aléatoire uniforme discrète X prenant des valeurs dans $\{1, 2, ..., m\}$.

(b) Calculez l'entropie d'une variable aléatoire de Bernoulli X avec le paramètre ϕ. Rappelons que si $X \sim Bernoulli(\phi)$ alors $p(x) = \phi^x(1-\phi)^{1-x}$, $x = 0, 1$.

(c) Pour quelle valeur ϕ^*, $H(X)$ est maximisé en fonction de ϕ?

(d) L'entropie croisée entre deux variables aléatoires discrètes X et Y est définie comme :

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

où $p(x, y) = P(X = x, Y = y)$.
Supposons que X et Y sont indépendants, c'est-à-dire $p(x, y) = p(x)p(y)$.
Montrez que $H(X, Y) = H(X) + H(Y)$.

3. Soit $Y \sim Bernoulli(\phi)$, et $D = y^{(1)}, ..., y^{(m)}$ un échantillon de observations identiquement indépendamment distribuées tirées de Y. (1.5 + 2.5 = 4 Points)

(a) Calculez la moyenne $E(Y)$ et la variance $V(Y)$.

(b) Calculez l'estimateur de maximum de vraisemblance de ϕ en maximisant $\ell(\phi) = \log\left(\prod_{i=1}^{m} p(y^{(i)})\right)$.

4. (1.5 + 1 + 1 + 1.5 = 5 Points)

(a) Soit $\sigma(x) = \frac{1}{1+e^{-x}}$, montrez que $\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$.

(b) Soient $\boldsymbol{x} = (x_1, x_2)^T \in \mathbb{R}^2$, $\boldsymbol{w} = (w_1, w_2)^T \in \mathbb{R}^2$ et $g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w} = \boldsymbol{x}.\boldsymbol{w}$.
Montrez que le gradient $\nabla_x g(\boldsymbol{x}) = \boldsymbol{w}$, où $\nabla_x g(\boldsymbol{x}) = (\frac{\partial g(\boldsymbol{x})}{\partial x_1}, \frac{\partial g(\boldsymbol{x})}{\partial x_2})^T$.

(c) Soient $\boldsymbol{x} = (x_1, x_2)^T \in \mathbb{R}^2$, $A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$, et $Q(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}$.
Calculez $\nabla_x Q(\boldsymbol{x})$ et comparez-le à $(A + A^T)\boldsymbol{x}$.

(d) En utilisant le résultat de (c) déduisez $\nabla_x \|\boldsymbol{x}\|_2^2$, où $\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{x}.\boldsymbol{x}$.