Corrigé

Examen ING2 EILCO - Ingénierie Mathématique

Décembre 2024

Nom :

Prénom :

Total: 32 points
Durée: 2h
Instructions générales: L'examen comprend 2 parties (Chacune de ces parties reprenant différentes sous-questions). Vous êtes libres de rédiger vos réponses sur des pages supplémentaires en veillant toutefois à bien indiquer le numéro de chaque question. Une fois l'examen terminé, Assurez vous de bien écrire votre nom (de façon lisible) sur chacune des pages. Répondez à un maximum de questions, en commençant par les questions qui vous semblent les plus abordables.

Question 1 (18pts)

1. [5*pts*] *Indiquer si les affirmations suivantes sont vraies ou fausses*

   Vrai / *Faux* — *En régression linéaire, l'estimateur de maximum de vraissemblance suppose que les données sont indépendantes et identiquement distribuées*

   Vrai / *Faux* — *Soit le noyau K défini par la matrice $K(x, y) = \tanh(\alpha(\boldsymbol{x}^\intercal \boldsymbol{y}) + \mathbb{1}^\intercal \boldsymbol{x})$ où $\mathbb{1}$ représente le vecteur $\mathbb{1} = [1, \ldots, 1]$ et* tanh *est la tangente hyperbolique. Le noyau K est un noyau valide.*

   Vrai / *Faux* — *Lorsque l'a priori est uniforme, l'estimateur de maximum a posteriori se reduit a un estimateur de maximum de vraissemblance.*

   Vrai / *Faux* — *Le modèle de régression logisitique nécessite que les caractéristiques suivent une distribution Gaussienne.*

   Vrai / *Faux* — *Dans l'algorithme du perceptron, la mise à jour est donnée par $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} + \eta t^{(i)} \tilde{\boldsymbol{x}}^{(i)}$ où les $t^{(i)}$ sont les valeurs cibles et les $\tilde{\boldsymbol{x}}^{(i)} = [1, \boldsymbol{x}^{(i)}]$ sont les vecteurs caractéristiques*

   Vrai / *Faux* — *Dans le cas du modèle de régression logistique, la fonction de vraissemblance peut prendre des valeurs négatives lorsque les vecteurs caractéristiques sont suffisamment différents*

   Vrai / *Faux* — *Un réseau de neurones entrainé via la mise à jour ADAM convergera toujours vers un minimum global de la fonction de coût*

2. [4*pts*] *Déterminer quelles sont, parmi les affirmations suivantes, celles qui sont correctes:*

   1) En scikit-learn, quelle fonction est utilisée pour diviser les données en un ensemble de test et un ensemble d'entrainement?

   A. data_split()
   B. fit_transform()
   C. split_data()

D. train_test_split()
E. cross_val_score()

2) *Que fait la fonction* StandardScaler() *en scikit-learn?*
   A. *Elle rééquilibre les caractéristiques de façon à ce que chaque caractéristique soit de moyenne nulle et de variance 1.*
   B. *Elle transforme les caractéristiques binaires en valeures réelles.*
   C. *Elle rééquilibre les caractéristiques de façon à ce que toutes ces caractéristiques soit comprises dans l'intervalle [0,1]*
   D. *Elle supprime les caractéristiques qui contiennent des valeurs* NaN *ou qui ne sont pas définies dans certaines données.*

3) *Un fois le modèle de régression linéaire entrainé en* scikit-learn, *quelle ligne permet de récupérer les coefficients du modèle?*
   A. model.coefficients_
   B. model.coef_
   C. model.weights_
   D. model.params_

4) *En* scikit-learn, *quelle est la classe utilisée pour entrainer un modèle de régression de type Ridge?*
   A. LinearRegression()
   B. RidgeRegression()
   C. Ridge()
   D. RidgeClassifier()
   E. Ridge_Regression()

5) *Quel paramètre du modèle* MLPClassifier *contrôle le nombre de neurones dans chaque couche?*
   A. n_neurons
   B. hidden_neurons
   C. hidden_layer
   D. hidden_layer_sizes
   E. network_size

6) *Quel est l'effet du paramètre* random_state *en scikit-learn?*
   A. *Il permet d'initialiser la descente de gradient de façon aléatoire*
   B. *Il permet de sélectionner de manière aléatoire un sous-ensemble d'entrainement parmi les données.*
   C. *Il permet de spécifier la fraction des données utilisée pour l'entrainement et pour le test*
   D. *Il permet d'assurer la reproductibilité des résultats en fixant le paramètre d'initialisation du générateur de nombres pseudo-aléatoires.*
   E. *Il permet d'ajouter une perturbation aléatoire aux données d'entrainement*

7) *Parmi les fonctions d'activation suivantes, quelle est celle qui n'est pas implémentée par le modèle* MLPClassifier*?*
   A. linear
   B. logistic
   C. relu
   D. heaviside
   E. tanh

3. [3*pts*] *On suppose qu'on dispose d'un vecteur caractéristique* $\boldsymbol{x}$ *pouvant appartenir soit à la classe* C_0 *soit à la classe* C_1 *(On suppose que les classes sont mutuellement exclusives). On défini les activations*

$$a_1 = \log P(x|M_1) + \log P(M_1) \tag{1}$$
$$a_2 = \log P(x|M_2) + \log P(M_2) \tag{2}$$

Montrer que la probabilité a posteriori du modèle M_1 peut se réécrire à l'aide de la fonction sigmoïde de la manière suivante:

$$P(M_1|x) = \sigma(a_1 - a_2) = \frac{1}{1 + \exp(-(a_1 - a_2))} \tag{3}$$

(Indice: Etant donné deux événements mutuellement exclusifs A et B, $p(x) = p(x|A)p(A) + p(x|B)p(B)$)

4. [4pts] *On considère un problème de classification à K classes.* → voir notes de cours

1/2 (a) *Donner les nombres de modèles linéaires devant être entrainés dans le cas de chacune des approches "un contre un" et "un contre tous".*

1/2 (b) *En quelques mots (sans donner le pseudo-code), expliquer la différence entre les deux approches.*

1/2 (c) *En quelques mots, expliquer comment ces deux approches peuvent conduire a une ambiguité quant à la classe de certaines régions de l'espace des caractéristiques.*

1/2 (d) *Finalement, donner une alternative permettant de résoudre cette ambiguité*

5. [2pts] *Donner l'expression de la fonction objectif dans le cas d'une régularisation de type Ridge.*

$$\ell(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(t^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)})\right)^2 + \lambda \sum_{j=1}^{D} \beta_j^2$$

Question 1.3 : Solution

En appliquant l'exponentielle à $(-a_1 + a_2)$ on trouve

$$\frac{1}{1 + \exp(-(a_1 - a_2))} = \frac{1}{1 + \exp(-\log P(x|M_1) - \log P(M_1) + \log P(x|M_2) + \log P(M_2))}$$

$$= \frac{1}{1 + \dfrac{P(x|M_2)P(M_2)}{P(x|M_1)P(M_1)}} = \frac{P(x|M_1)P(M_1)}{P(x|M_1)P(M_1) + P(x|M_2)P(M_2)}$$

on conclue en utilisant le fait que les classes M_1 et M_2 sont mutuellement exclusives $P(x|M_1)P(M_1) + P(x|M_2)P(M_2) = P(x)$

ce qui donne

$$\frac{1}{1 + \exp(-(a_1 - a_2))} = \frac{P(x|M_1)P(M_1)}{P(x)} = \frac{P(x, M_1)}{P(x)} = P(M_1|x)$$

Question 2 (14pts)

1. [5pts] *Indiquer si les affirmations suivantes sont vraies ou fausses*

Vrai / Faux	*L'algorithme de rétropropagation ne fonctionne que pour des réseaux de neurones dont les couches sont complètement connectées*
Vrai / Faux	*Une famille de modèles dont l'erreur sur les données d'entrainement est nulle est nécessairement associée à une variance faible.*
Vrai / Faux	*En régression logistique, les points abhérents peuvent avoir un effet important sur l'estimateur des coefficients de régression*
Vrai / Faux	*Les modèles Ridge et Lasso supposent tous les deux que les erreurs entre les données exactes et le modèle linéaire sous-jacent suivent une loi Gaussienne*
Vrai / Faux	*Un niveau de régularisation élevé a tendance à augmenter le biais de la famille de modèles associés.*
Vrai / Faux	*L'astuce du noyau est particulièrement utile pour entrainer des modèles sur des jeux de données de grande taille.*
Vrai / Faux	*En validation croisée à K compartiments, les données sont séparées en K sous-ensembles et chaque sous-ensemble est utilisé une fois comme ensemble de test, tandis que les $K-1$ sous-ensembles restants sont utlisés comme ensembles d'entrainement.*
Vrai / Faux	*Si un jeu de données est linéairement séparable par un modèle de régression logistique de la forme $\sigma(\boldsymbol{w}^T\boldsymbol{x}+b)$ ou $\sigma(a)=1/(1+e^{-a})$, alors il reste linéairement séparable pour tout modèle de la forme $\sigma(\boldsymbol{w}_c^T\boldsymbol{x}+b_c)$ où $\boldsymbol{w}_c=c\boldsymbol{w}$ et $b_c=cb$.*

2. [4pts] *On considère l'extrait de code donné à la figure 1. Compléter la fonction* my_linear_regression() *de façon à satisfaire les spécifications. On supposera que l'ordinateur ne dispose pas de la librairie* scikit-learn. *On veillera à bien détailler chaque étape.*

3. [5pts] *On considère le problème de classification suivant. On dispose de vecteurs caractéristiques constitués de deux caractéristiques x_1 et x_2, ainsi que d'une valeur cible binaire $t \in \{0,1\}$. Pour chaque vecteur caractéristique, la valeur cible est définie de la manière suivante:*

$$t(\boldsymbol{x}) = \begin{cases} 1 & si\ x_2 \geq |x_1| \\ 0 & sinon \end{cases} \tag{4}$$

   (a) [1pt] *La fonction $t(\boldsymbol{x})$ peut-elle être représentée par un modèle de régression logistique ou par un perceptron? Justifier.*

   (b) [2pts] *On considère une simplification de la fonction (4) donnée par*

$$t(\boldsymbol{x}) = \begin{cases} 1 & si\ x_2 \geq x_1 \\ 0 & sinon. \end{cases} \tag{5}$$

   Déterminer l'expression ($\boldsymbol{w}=(w_1,w_2)$ et b) d'un modèle de perceptron de la forme $\sigma(\boldsymbol{w}^T\boldsymbol{x}+b)$, ou $\sigma(a)$ est la fonction d'activation

$$\sigma(a) = \begin{cases} 1 & si\ a \geq 0 \\ 0 & sinon \end{cases}, \tag{6}$$

   permettrant de représenter $t(\boldsymbol{x})$,

   (c) [2pts] *On souhaite à présent revenir aux cibles introduites en (4). Donner l'expression $(\alpha_1, \alpha_2, \alpha_0, \boldsymbol{w}, \tilde{\boldsymbol{w}}, b, \tilde{b})$ d'un réseau de neurones a deux couches, i.e.*

$$y(\boldsymbol{x}) = \sigma(\alpha_1\sigma(\boldsymbol{w}^T\boldsymbol{x}+b) + \alpha_2\sigma(\tilde{\boldsymbol{w}}^T\boldsymbol{x}+\tilde{b}) + \alpha_0) \tag{7}$$

In [ ]:

```python
import numpy as np
import matplotlib.pyplot as plt


def my_linear_regression(X, t, eta):

    '''la fonction doit renvoyer le vecteur des coefficients de régression d'un
    modèle entrainé sur base des données X, t'''

    # Entrées:    X : Matrice caractéristique.
    #             t : vecteur des valeurs cibles
    #           eta : taux d'apprentissage
    # Sortie:  beta : Vecteur des coefficients de régression


    '''À compléter'''




    return beta

```

Figure 1: Extrait de code utilisé à la question 2.2.

permettant de représenter la fonction. Indice: On pourra ré-utiliser l'expression du neurone calculée en (3b) *pour l'un des neurones de la couche cachée.*

Solution Question 2.3

3a : Non, les données ne sont pas linéairement séparables

3b : Il suffit de prendre $y(x) = \sigma(x_2 - x_1)$

3c : Une possibilité consiste à utiliser $y_1(x) = \sigma(x_2 - x_1)$ pour le neurone 1 et $y_2(x) = \sigma(x_2 + x_1)$ pour le neurone 2 de la première couche.

On a alors $y_1(x) + y_2(x) = \begin{cases} 2 & x_2 \geq |x_1| \\ 1 & x_2 < x_1 \text{ et } x_2 > -x_1 \\ & \text{ou } x_2 > x_1 \text{ et } x_2 < -x_1 \\ 0 & x_2 < -|x_1| \end{cases}$

On peut donc prendre $y(x) = \sigma(y_1(x) + y_2(x) - 2)$

ce qui donne 1 si $x_2 \geq |x_1|$ et 0 autrement.