

Université de Montréal
Data Science
IFT-6758 – Examen Intra

2022/11/01

Nom Complet (Prénom Nom)

Numéro de matricule (UdeM)

--	--	--	--	--	--	--	--

Instructions

- Aucune ressource externe n'est autorisée.
- Donnez de brèves explications pour vos réponses. Nous accorderons un crédit partiel pour un bon raisonnement.
- Pour les questions avec une boîte à remplir, indiquer votre réponse dans la boîte.
- Il y a du papier supplémentaire à l'avant de la pièce. Assurez-vous de mettre votre nom, votre identifiant et votre numéro de question pour toutes les pages supplémentaires que vous utilisez.

Notations and rappels:

- $\|x_i - x_j\|_2^2 = \sum_{k=1}^d ([x_i]_k - [x_j]_k)^2$ = Euclidean distance between points x_i and x_j
- $A \setminus B$ = Complémentaire d'un ensemble: éléments dans A mais pas dans B
- $H(X) = -\sum_{i=1}^d p(X = x_i) \log(X = x_i)$, entropie d'un variable aléatoire discrète X .
- $I(X, Y) = \sum_{i,j} p(X = x_i, Y = y_j) \log \left(\frac{p(X=x_i, Y=y_j)}{p(X=x_i)p(Y=y_j)} \right)$ information mutuelle (ou gain d'information) entre deux variables aléatoires X et Y .
- Taux de vrai positifs : $= \frac{\#(\text{prédiction}=+|\text{vrai valeur}=+)}{\#(\text{vrai valeur}=+)}$, Taux de vrai négatifs: $= \frac{\#(\text{prédiction}=-|\text{vrai valeur}=-)}{\#(\text{vrai valeur}=-)}$

Distribution des points

Question	Points	Score
1	2	
2	12	
3	13	
4	10	
5	14	
6	10	
7	13	
8	6	
9	10	
Total:	90	

1. Échauffement

- (a) (1 point) Écrire correctement et lisiblement votre nom complet et numéro de matricule (voir Figure 1).
- (b) (1 point) (À la fin) Mettre sa copie dans la bonne pile (marquée par la **première lettre** de votre **prénom**)

Full Name (Firstname LASTNAME)

Harry James POTTER

Student ID (UdeM matricule)

	3	1	7	1	9	8	0
--	---	---	---	---	---	---	---

Instructions

Figure 1: ou comment marquer un point facilement.

2. **Puissance d'un test.** Le média social X utilise un test T pour détecter les publications qui enfreignent les directives communautaires de la plateforme. Nous supposons que 1000 messages de *comptes non suspects* (c'est à dire jamais signalé auparavant) ont été créés sur cette plateforme, que le classificateur a un taux de vrais positifs et un taux de vrais négatifs de 90 %, et que parmi les messages, 10% enfreignent les directives communautaires de la plateforme (qualifiées de "vrais positifs").

	prédit positif	prédit négatif	Total
Vrai positif	TP: 90	FN: 10	100
Vrai négatif	FP: 90	TN: 810	900

Table 1: Nombre de publication dans chaque situation.

- (a) (2 points) Complétez les quatre cellules du tableau 1.

J'ai ajouté une colonne qui aide à remplir le tableau et j'ai étiqueté chaque cellule. TP : vrai positif, FN : faux négatif, FP : faux positif, TN : vrai négatif.

- 10% des messages enfreignent les règles de la communauté : 100 vrai positifs.
 - 90% taux de vrais positifs : 90 prédictions positives et 10 prédictions négatives (parmi la première ligne).
- 10% des messages enfreignent les directives de la communauté : 900 vrai négatifs.
 - 90% taux de vrais négatifs : 810 prédictions négatives et 90 prédictions positives (parmi la première ligne).

- (b) (2 points) Une publication a été signalée (prédite positive) par le classificateur. Elle vient d'un utilisateur qui ***n'a jamais été signalé par la plate-forme auparavant***. Quelle est la probabilité que la publication enfreigne réellement les directives de la communauté ?

RÉPONSE (sous forme de fraction): (expliquez pourquoi)

Nous voulons calculer $\frac{TP}{TP+FP} = \frac{90}{90+90} = \frac{1}{2}$ (c'est-à-dire la probabilité d'être un réel positif étant donné qu'il a été prédit positif)

- (c) (2 points) Pour la publication mentionnée dans la question b) : que dois-je faire ?

La supprimer : ☐ La faire réviser par un humain : ☐ Ne rien faire : ☐

(expliquez pourquoi)

Dans ce cas, la meilleure réponse était de **le faire réviser par un humain**.

Pourquoi? Parce que le nombre de messages positifs prédits n'est pas si grand (180 sur 1000) donc le coût à faire réviser par un humain est raisonnable, et seulement la moitié d'entre eux sont réellement positifs (nous devons donc nous assurer que c'est un vrai positif avant de le retirer).

La suppression du message conduirait à supprimer trop de messages qui ne sont pas réellement positifs. Ne rien faire conduirait à accepter tous les messages qui violent les directives.

- (d) (2 points) Une publication **n'a pas été signalée** par le classificateur (prédite négative). Elle vient d'un utilisateur qui **n'a jamais été signalé** par la plate-forme auparavant. Quelle est la probabilité que la publication enfreigne les règles de la communauté?

RÉPONSE (sous forme de fraction): (expliquez pourquoi)

Nous voulons calculer $\frac{FN}{TN+FN} = \frac{10}{10+810} = \frac{1}{82}$ (c'est-à-dire la probabilité d'être un réel positif étant donné qu'il a été prédit négatif)

- (e) (2 points) Pour la publication mentionnée dans la question d) : que dois-je faire ?

La supprimer : La faire réviser par un humain : Ne rien faire :
(expliquez pourquoi)

Dans ce cas, nous ne faisons rien. L'examen de tous les messages qui sont prédits négatifs est trop coûteux et la probabilité que le message soit réellement positif est faible (1/82).

- (f) (2 points) Une publication a été signalée (prédite positive) par le classificateur. Elle vient d'un utilisateur qui **a déjà été signalé** par la plate-forme. Que devrais-je faire?

La supprimer : la faire réviser par un humain : Ne rien faire :
(expliquez pourquoi)

Le fait que l'utilisateur ait été signalé auparavant lui donne plus de chance de publier un vrai positif. Cela ne correspond donc pas aux données du tableau 1 qui proviennent de *comptes non suspects*. Ainsi, dans ce cas, la suppression du post semble être l'action la plus efficace (sans frais). Cependant, en fonction du budget et de la probabilité que le poste soit un réel positif, nous pourrions affirmer que nous voulons toujours le faire réviser.

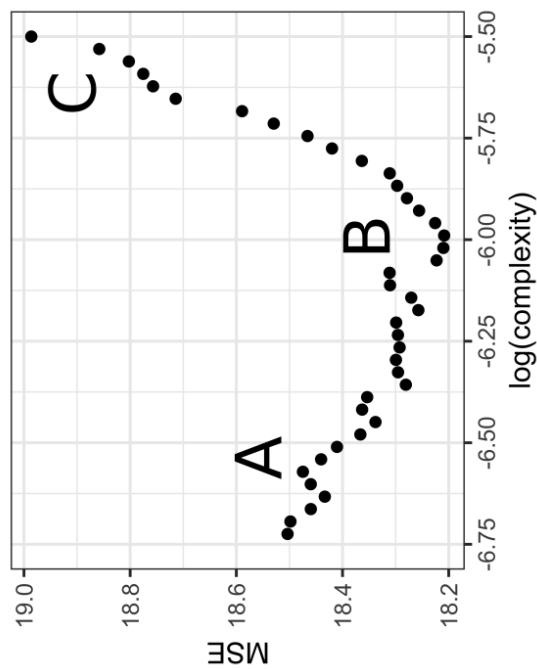
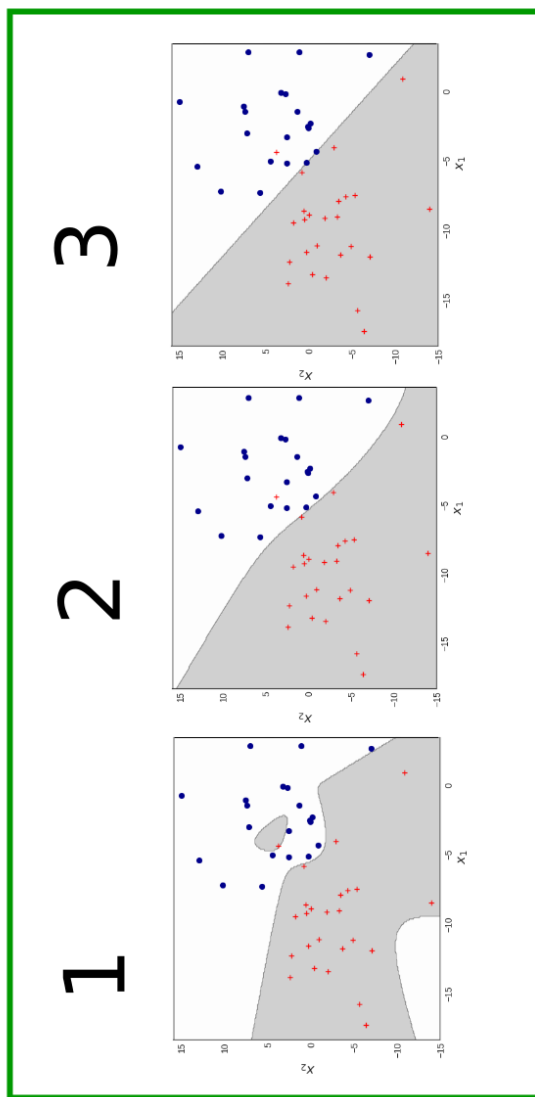


Figure 2: Figure pour la question 3.

3. **Complexité des modèles.** Premièrement, nous considérons une tâche de classification apprise par régression logistique avec des caractéristiques polynomiales. Plus précisément, étant donné un vecteur de caractéristiques bidimensionnel (x_1, x_2) nous considérons les prédicteurs :

$$f_1(x_1, x_2|w) = w_0 + w_1x_1 + w_2x_2$$

$$f_2(x_1, x_2|w) = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2 + w_6x_1^3 + w_7x_2^3 + w_8x_1^2x_2 + w_9x_1x_2^2$$

$$f_3(x_1, x_2|w) = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2$$

$$f_4(x_1, x_2|w) = w_1x_1 + w_2x_2 + w_3 \cdot (x_1 + x_2) + w_4 \cdot (x_1 - x_2)$$

$$f_5(x_1, x_2|w) = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2$$

$$f_6(x_1, x_2|w) = w_0 + w_1x_1 + \dots + w_4x_1^4 + w_5x_2 + \dots + w_8x_2^4 + w_9x_1x_2 + w_{10}x_1x_2^2 + w_{11}x_1x_2^3 \\ + w_{12}x_1^2x_2 + w_{13}x_1^2x_2^2 + w_{14}x_1^3x_2$$

- (a) (4 points) Trier les 6 modèles f_1, \dots, f_5, f_6 par complexité :

(Élevé) \gg ; \gg \gg ; \gg ; \gg (Faible). Expliquer:

La réponse était $f_6 \gg f_2 \gg f_5 \gg f_3 \gg f_1 \gg f_4$.

Tous ces modèles correspondent à des modèles linéaires à caractéristiques polynomiales. La complexité du modèle dépend donc du degré du polynôme.

f_6 : polynôme de degré 4 avec interactions

f_2 : polynôme de degré 3 avec interactions

f_5 : polynôme de degré 2 avec interactions

f_3 : polynôme de degré 2 avec **aucune** interaction (pas de x_1x_2)

f_1 : polynôme de degré 1, aussi appelé fonction affine

f_4 : peut être réécrit comme $f_4(x_1, x_2|w) = (w_4 + w_3 + w_1)x_1 + (w_2 + w_3 = w_4)x_2$. C'est donc une fonction linéaire (pas de partie affine).

- (b) (3 points) Dans le panneau vert de la Figure 2, nous avons tracé les limites de décision de f_1, f_3 et f_6 . Quel tracé correspond à quelle fonction ? (justifier)

Une frontière de décision plus complexe correspond à un polynôme de degré supérieur.

Tracé 1 : correspond à la frontière de décision la plus complexe : c'est f_6 .

Tracé 2 : correspond à une frontière de décision non linéaire moins complexe que le tracé 1 : c'est f_3 .

Tracé 3 : correspond à une frontière de décision affine : c'est f_1

Arbres de décision : En dehors du panneau vert (panneau supérieur et en bas à droite) de la figure 2, nous présentons trois arbres de décision.

- (c) (2 points) Comment contrôler la complexité des arbres de décision ?

En contrôlant la profondeur de l'arbre, en enlevant (pruning) quelques branches. On peut utiliser des forêts aléatoires. On peut limiter le nombre de feuilles.

- (d) (4 points) Le coin inférieur gauche de la Figure 2 montre trois points A, B, C sur la courbe biais-variance pour une collection de modèles d'arbre de décision. Pour chaque arbre (panneau supérieur), remplissez la lettre du point correspondant dans la courbe biais-variance. Justifier brièvement.

1:	A
2:	C
3:	B

Le premier arbre est le moins complexe (sur la figure, l'axe des abscisses correspond à la complexité de l'arbre). Ainsi, cela correspond à la partie sous-entraîné du tracé : la fonction perte diminue à mesure que nous augmentons la complexité.

Le deuxième arbre est le plus complexe. Ainsi, cela correspond à la partie de sur-entraînée du tracé : la fonction de perte augmente à mesure que nous augmentons la complexité (la frontière de décision est trop complexe et ne se généralise pas)

Le troisième arbre se situe entre les deux : il correspond donc au meilleures performances : lorsque la fonction de perte est au plus bas.

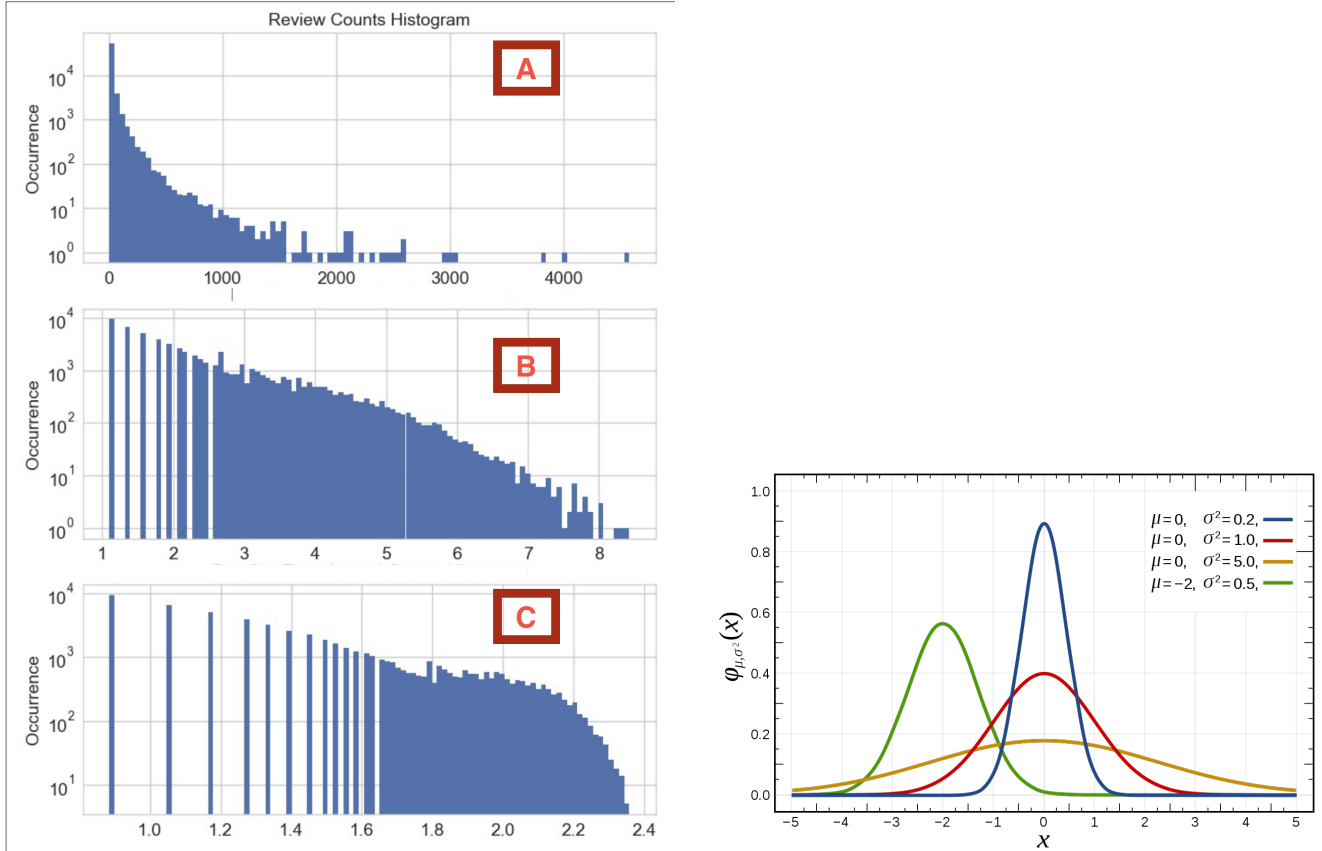


Figure 3: **Gauche:** Histogrammes des valeurs transformées des décomptes des critiques pour trois valeurs différentes de λ (reviews). **Droite:** Densité Gaussiennes $\varphi_{\mu, \sigma^2}(x)$ pour différentes valeurs de la moyenne μ et de la variance σ .

4. Ingénierie de Caractéristiques

La transformation de Box-Cox est définie pour $\lambda \in \mathbb{R}$ comme suit (rappel : $e^{\lambda \log(x)} = x^\lambda$):

$$T_\lambda(x) = \begin{cases} \frac{e^{\lambda \log(x)} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0. \end{cases} \quad (1)$$

- (a) (2 points) Quelle est l'interprétation de la transformation de Box-Cox pour $\lambda = 0, \frac{1}{2}, 1$?

$\lambda = 0$: il s'agit d'une simple transformation logarithmique.

$\lambda = 1$: $T_1(x) = x - 1$ c'est un décalage de 1.

$\lambda = 1/2$: $T_{1/2}(x) = 2\sqrt{x} - 2$. C'est une transformation de racine carrée.

(b) (2 points) La transformation de Box-Cox fonctionne seulement pour:

des données positives ($x > 0$): des données négatives ($x < 0$):
 des données non-nulles $x \neq 0$: tout type de données:

Expliquer que faire si l'on a des caractéristiques qui ne sont pas dans le bon ensemble:

Prendre la valeur absolue peut ne pas être une bonne idée si les valeurs positives et négatives ont des significations radicalement différentes (par exemple, si x est le retard pour le départ d'un avion, une valeur négative correspond à un avion qui part tôt (cela se produit si tous les passagers sont déjà en l'avion) alors qu'une valeur positive correspondrait à un avion partant en retard).

En fonction de l'application, vous souhaitez peut-être :

- Coupe les valeurs (les valeurs inférieures à $\epsilon > 0$ sont coupées à $\epsilon > 0$)
- Supprimez le minimum des valeurs observées et ajoutez ϵ .
- Dans certains cas, $\epsilon = 1$ est un bon choix.

(c) (2 points) On veut trouver la valeur de λ pour laquelle la transformation de la distribution est la plus proche d'une distribution Gaussienne (voir Figure 3). Décrire une méthode pour choisir λ ?

Habituellement, un tel λ est trouvé en maximisant la probabilité d'être gaussien. (maximum log likelihood estimation)

Cette question était délibérément ouverte. Toute réponse décrivant que l'on peut utiliser une mesure de normalité pour rechercher un λ qui maximise cette métrique obtiendrait les points.

Le calcul de l'histogramme n'est pas suffisant car il ne fournit pas de métrique quantitative.

- (d) (2 points) Dans la Figure 3 on représente les histogrammes des valeurs transformées des décomptes des critiques (reviews). Quelle est la figure qui correspond à la meilleure approximation Gaussienne? A: ☐ B: ☐ C: ☐ (Justifier)

A correspond à la distribution des données brute.

B correspond à une transformation logarithmique.

C correspond à la meilleure valeur pour λ .

Le côté gauche de la distribution ne ressemble jamais à une gaussienne (voir question suivante). C'est l'histogramme C qui ressemble le plus à une gaussienne car il a une forme de cloche. En particulier, pour la distribution gaussienne, les échantillons éloignés de la moyenne sont très peu probables (les tracés A et B ont des exemples assez éloignés de la moyenne)

- (e) (2 points) En utilisant le fait que le nombre de critiques est toujours positif, expliquer le fait que on ne peut pas s'attendre à ce que l'approximation Gaussienne soit parfaite sur la partie gauche de l'histogramme des la distribution après transformation ($x < 0$ pour A, $x < 2.5$ pour B, $x < 1.6$ pour C).àà

L'idée ici était que nous avons un ensemble de données avec uniquement des valeurs positives. Parce qu'une distribution gaussienne donne des **probabilités non nulles** aux valeurs négatives, il est impossible d'avoir un ajustement gaussien parfait.

5. **Sélection des Caractéristiques:** Votre équipe de science des données utilise un ensemble de données comme illustré dans la Fig. 2 pour prédire l'espérance de vie (en années) des patients en fonction de certains facteurs mesurables. Détail sur les caractéristiques : **ID** : numéro d'identification unique anonyme. **imc** : indice de masse corporelle calculé comme le rapport entre la masse (en kg) et la taille au carré (en m^2). **grai_corp** : proportion de graisse dans le corps. **freq_card** : fréquence cardiaque moyenne (sur une journée) de la personne en nombre

de battements par minute. **glycémie** : quantité de sucre (en mg) par litre de sang (en L).
espérance_vie : âge auquel le patient est décédé.

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	y
ID	sexe (à la naissance)	taille (cm)	taille (ft)	poids (kg)	poids (lb)	imc (kg/m ²)	grai_corp (%)	freq_card (bpm)	glycémie (mg/L)	espérance_vie (years)
1012	M	154	5.23	48	105.82	20.2	4	68	15	68
2543	F	162	5.51	65	143.3	24.8	13	70	11	73
3674	F	177	6.02	78	171.96	24.9	20	66	14	89
4435	M	184	6.26	84	185.19	24.8	20	80	18	88
5809	F	171	5.81	67	147.71	22.9	31	50	21	59

Table 2: Caractéristiques et Étiquettes.

- (a) (2 points) Énumérez les caractéristiques qui vous semblent *non pertinentes*. **Justifiez** vos choix.

La fonctionnalité non pertinente est f_1 . Les autres fonctionnalités ne sont **pas** non pertinentes car elles pourraient être utiles pour la classification.

Certains d'entre vous ont fait valoir que f_9 et f_{10} pouvaient, dans une certaine mesure, ne pas être pertinents. Il ne fallait néanmoins pas oublier de mentionner que f_1 est la fonctionnalité la moins pertinente.

- (b) (2 points) Listez les caractéristiques que vous pensez être *redondantes*. **Justifiez** vos choix.

Les fonctionnalités redondantes sont : **taille**, **poids** et **imc**. L'IMC peut être dérivé du poids et de la taille. La taille et le poids dans différentes unités sont redondants (car ils sont proportionnels).

- (c) (3 points) Votre équipe est divisée entre l'utilisation des méthodes d'encapsulation et de filtrage pour la sélection des caractéristiques. Indiquez **un argument chacun** en faveur de la position de chacun de ces deux camps.

Wrapper (encapsulation) : plus coûteux en termes de calcul, mais obtiendrait les meilleures performances sur un algorithme d'apprentissage automatique donné.

Filtre : moins cher en termes de calcul, avec des performances potentiellement moins bonnes avec un modèle ML donné, mais les fonctionnalités sélectionnées pourraient être généralisées à plusieurs modèles ML.

De plus, les méthodes de filtrage sont généralement plus interprétables.

- (d) (2 points) Indiquez une méthode de filtrage spécifique qui peut supprimer à la fois les caractéristiques non pertinentes et redondantes que vous avez identifiées dans (a) et (b).

Mentionner les méthodes de filtrage basées sur la théorie de l'information (information mutuelle, gain d'information) était la réponse attendue.

- (e) (2 points) Comment utiliser LASSO pour éliminer les caractéristiques que vous avez jugées non pertinentes et redondantes et dans (a) et (b) ?

La méthode LASSO utilise une pénalité ℓ_1 pour régulariser la solution. La pénalité ℓ_1 incite à la parcimonie de la solution. Les caractéristiques non pertinentes et redondantes sont les caractéristiques qui seraient éliminées en premier lors de la recherche d'une solution parcimonieuse car elles n'influencent pas/peu la prédiction.

- (f) (3 points) Supposons que les entités puissent être considérées comme des groupes - $G_1 = \{f_3, f_4\}$, $G_2 = \{f_5, f_6\}$, $G_3 = \{f_1, f_2, f_7, \dots, f_{10}\}$. Laquelle des trois méthodes : LASSO, Group LASSO et Sparse Group LASSO est idéale pour la sélection des caractéristiques ? Donnez un ensemble possible (valide) de caractéristiques sélectionnées par ces modèles en complétant le Tableau 3 (Cochez les cases des caractéristiques sélectionnées et justifiez sous le tableau.)

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}

Table 3: Cochez les cases des caractéristiques sélectionnées

La réponse attendue était Sparse Group LASSO. Nous ne voulons pas utiliser tous les groupes et à l'intérieur de chaque groupe, nous ne voulons utiliser que certaines fonctionnalités (pour éviter la redondance).

Les personnes qui ont répondu groupe LASSO et LASSO ont obtenu des points au cas où elles proposeraient une sélection cohérente sur les fonctionnalités. (Dans le cas du groupe LASSO un groupe est activé ou désactivé)

6. **(Code)** Cinq fonctions mystères (étiquetées A-E) implémentées en Python sont données ci-dessous. Associez chaque description à l'extrait de code approprié. **Justifiez votre réponse.**

```
1 def fonction_mystere_A(d, c):
2     return c*numpy.sum(numpy.abs(d))
3
4 def fonction_mystere_B(d, c=r"(\ )+"):
5     for t in d:
6         if re.match(c, t):
7             return True
8     return False
9
10 def fonction_mystere_C(d, c):
11     n = 0
12     for g in c:
13         n += numpy.sqrt(numpy.sum(d[g]**2))
14     return n
15
16 def fonction_mystere_D(d, c):
17     r = numpy.sum((d-c)**2)
18     t = numpy.sum((d-numpy.mean(c))**2)
19     return 1 - r/t
20
21 def fonction_mystere_E(d, c):
22     return numpy.sum(numpy.log(d/(c["x"]*c["y"]).T) * d)
```

Assurez-vous de **justifier** votre choix.

- (a) (2 points) Calculer le coefficient de détermination R^2 étant donné les prédictions et les vraies valeurs.

fonction_mystere_

D

Calcule une erreur quadratique moyenne entre c et d.

- (b) (2 points) Renvoie si une expression régulière spécifique est trouvée dans l'un des textes d'un corpus (liste de textes).

`fonction_mystere`

B

utilise la fonction `re.match`. A une regex dans ses arguments.

- (c) (2 points) Calcule la pénalité LASSO d'un vecteur donné.

`fonction_mystere`

A

calcule la sommes des valeurs absolues

- (d) (2 points) Calculer l'information mutuelle (ou gain d'information) entre deux distributions

`fonction_mystere`

E

La seule fonction avec un logarithme

- (e) (2 points) Étant donné une liste de groupes, calculez la pénalité LASSO du groupe.

`fonction_mystere`

C

Somme sur les groupes g . Calcule une somme de normes ℓ_2 .

7. Statistiques Inférentielles

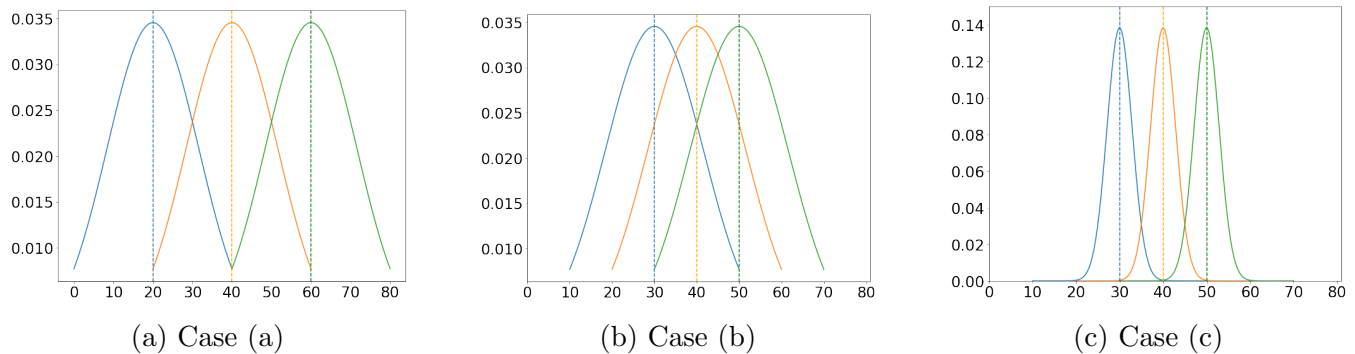


Figure 4: Trois distributions empiriques

Chaque figure représente les répartitions empiriques des échantillons des trois groupes différents dans les données (comme indiqué par les trois couleurs différentes). Les lignes verticales en pointillés représentent les moyennes. Notez la plage différente de l'axe des ordonnées pour le tracé (c).

- (a) (4 points) Comment pouvons-nous vérifier si les moyennes de ces trois groupes sont significativement différentes compte tenu de ce que nous savons des données dans chacun de ces cas (a), (b) et (c) ? En particulier qu'elle est l'hypothèse nulle? Quel est le test que l'on va utiliser et comment va-t-on procéder?

Solution : Nous devrions utiliser le test t pour comparer chaque groupe et **recalculer l'ajustement de la valeur p pour plusieurs classes**. Ou nous pouvons utiliser une analyse ANOVA.

- (b) (2 points) Dans quel cas les moyennes empiriques (marquées par des pointillés) sont les plus éloignées ? **Pourquoi ?**

(a) ☐ X (b) ☐ (c) ☐

- (c) (2 points) Dans quel cas la différence entre les moyennes empiriques semble la plus significative ? (a) ☐ (b) ☐ (c) ☒ X

parce que les chevauchements entre les distributions sont plus petits.

- (d) (2 points) Qu'est-ce que le théorème central limite ?

Si X_1, X_2, \dots, X_n sont des échantillons aléatoires tirés d'une population avec une moyenne globale μ et une variance finie σ^2 , et si \bar{X}_n est la moyenne des premiers n échantillons, alors la limite de la distribution, $Z = \lim_{n \rightarrow \infty} \left(\frac{\bar{X}_n - \mu}{\sigma_{\bar{X}}} \right)$, avec $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, est une distribution normale standard.

En clair : à la limite de l'échantillonnage infini, la **moyenne** des échantillons se comporte comme une gaussienne.

- (e) (3 points) Comment pouvons-nous utiliser le théorème central limite (ou n'importe quel test par paires) pour pirater les p-values ? àà

Nous pouvons l'utiliser pour tenter le test plusieurs fois. À la limite, le test positif ou négatif tendra vers la quantité appropriée, mais il y aura des tests qui seraient au-dessus ou en dessous de la valeur p finale.

Fondamentalement, ajouter des échantillons et s'arrêter si l'on obtient le résultat souhaité, c'est du p-hacking. Tester plusieurs hypothèses sur les mêmes données et ne signaler qu'une hypothèse sélectionnée est également du p-hacking.

8. **Test d'hypothèse.** Vous avez un ensemble de données D qui comprend les notes de tous les élèves de la province de Québec.

1. Hypothèse : IFT-6758 est la classe la plus difficile
2. Analyse : nous avons examiné toutes les notes

Fournissez 2-3 phrases pour chacune des questions suivantes.

- (a) (2 points) Comment modifieriez-vous les hypothèses pour qu'elles soient plus claires et utiles ?

Solutions : Définir le plus difficile, puis définir la portée comme étant la province de Québec. Les étudiants peuvent également noter que cette hypothèse devrait être temporelle (IFT6758 est la classe la plus difficile en ce moment)

- (b) (2 points) Comment modifieriez-vous l'analyse pour qu'elle soit plus reproductible (et compréhensible) ?

Solution : Donnez une description de la manière dont ils ont comparé les classes. Faut-il comparer IFT6758 aux classes de lycée ? Définissez "regardé" ce qui est regardé, pourquoi est-ce une mesure de difficulté ?

- (c) (2 points) Quels détails supplémentaires sur l'ensemble de données faut-il fournir pour améliorer l'analyse et la reproductibilité ? *Indice*: N'importe quel ensemble de données ne suffira pas si nous voulons avoir la réponse la plus précise.

Solution : Recueillir des données d'années supplémentaires. Notez que l'ensemble de données est une collection des données les plus récentes de la province et est à jour.

9. Virtualisation et partage de code:

- (a) (2 points) Quelle est la propriété la plus importante pour le succès de la science et aussi pour votre propre travail ?

Reproductibilité. Les résultats sont utiles, mais seulement s'ils sont cohérents et s'ils peuvent être réutilisés et améliorés.

- (b) (2 points) Vous avez évalué votre code sur un nouveau modèle qui s'entraîne sur l'ensemble de données D . Quelqu'un de votre équipe souhaite écrire des tests unitaires pour votre code, à quel niveau devez-vous partager votre environnement d'entraînement?

Partager le code: ☐ CONDA: ☐ Docker ☐

Partager le code ou Docker : il est peut-être possible de tester le code avec uniquement le code source. Les détails des tests n'ont pas été fournis.

- (c) (2 points) Vous avez évalué votre code sur un nouveau modèle qui s'entraîne sur l'ensemble de données D . Un membre de votre équipe souhaite utiliser votre algorithme d'apprentissage sur un nouvel ensemble de données D_1 . À quel niveau devez-vous partager votre environnement d'entraînement ?

Partager le code: ☐ CONDA: ☐ Docker ☐

Conda ou Docker : parce que nous devons réutiliser le même code et que nous souhaitons également utiliser leur propre ensemble de données.

- (d) (2 points) Vous avez évalué votre code sur un nouveau modèle qui s'entraîne sur l'ensemble de données D . Votre équipe est maintenant prête à déployer votre modèle sur le serveur pour une utilisation avec les clients, à quel niveau devez-vous partager votre environnement d'entraînement ?

Partager le code: ☐ CONDA: ☐ Docker ☐ Docker : Docker permet

de rendre les résultats indépendants de la configuration matérielle.

(e) (2 points) Quelle est la différence entre une machine virtuelle et un conteneur

La principale différence entre les conteneurs et les machines virtuelles est que les machines virtuelles virtualisent une machine entière jusqu'aux hardware et que les conteneurs ne virtualisent que les couches logicielles au-dessus du niveau du système d'exploitation.

Une machine virtuelle est un "ordinateur" à part entière (même le matériel est émulé) qui s'exécute à l'intérieur de votre ordinateur existant, tandis qu'un conteneur est un environnement léger et isolé qui s'exécute dans le système d'exploitation de l'ordinateur hôte.

Page supplémentaire 1 (si nécessaire)

Utilisez clairement les numéros de question et de partie si cette page doit être prise en compte

Page supplémentaire 2 (si nécessaire)

Utilisez clairement les numéros de question et de partie si cette page doit être prise en compte