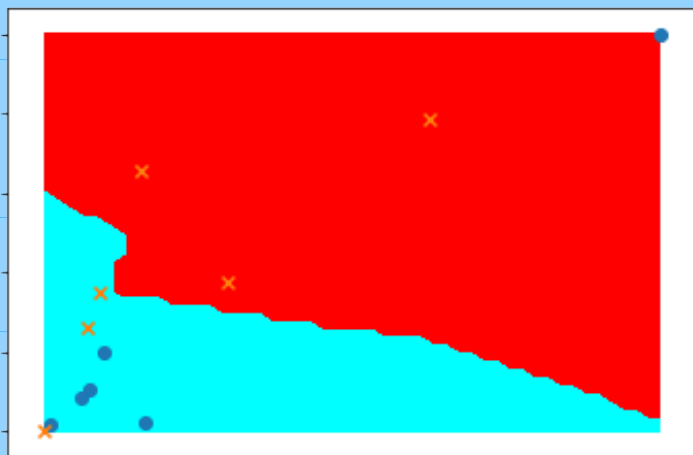
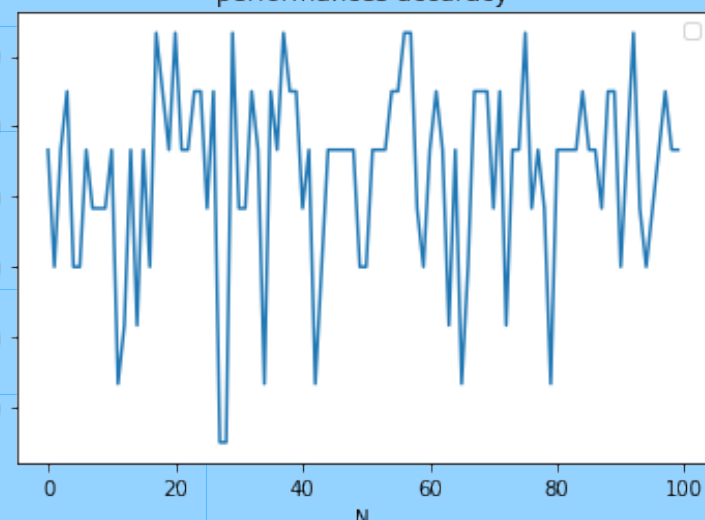


Introduction à l'IA et au traitement de grandes masses de données

Ryan OHOUENS		Sarah SALEH					
<div><div><div>Problématique 1 (Echauffement)</div><div><div>Description :</div><div>Nous souhaitons prédire si une catégorie sera rentable à partir des anciens films récoltés.</div></div><div><div>Méthode</div><div><div>Nous allons utilisé la médiane pour attribuer une classe à nos données.</div><div>Puis nous allons utiliser de la classification supervisée avec l’algorithme de k plus proches voisins.</div></div></div><div><div>Résultats :</div><div><p>Frontière de décision avec l'algorithme de KNN</p></div></div></div></div>		<div><div><div>Problématique 2</div><div><div>Description :</div><div>Nous souhaitons maintenant savoir quels sont les bons ingrédients pour la réalisation d'un film, soit comment obtenir la meilleure note possible.</div></div><div><div>Méthode</div><div><div>Nous allons toujours utilisé la méthode de la médiane pour attribuer une classe à nos données.</div><div>Ensuite, nous utilisons une régression supervisée grâce à l'algorithme des moindres carrés.</div><div>Puis nous nous entrainerons avec différentes combinaisons de paramètres pour savoir quels critères il faut prendre en compte pour la réalisation d'un film.</div></div></div><div><div>Résultats :</div></div></div></div>		<div><div><div>Problématique 3</div><div><div>Description :</div><div>Nous désirons désormais déterminer la catégorie d'un film à partir des caractéristiques.</div></div><div><div>Methode :</div><div><div>Nous utilisons toujours la méthode de la médiane pour attribuer une classe à nos données.</div><div>Par la suite, nous utilisons de la classification multi-classes. Pour cela, nous allons créer 20 data frames qui emploieront la méthode un contre tous pour permettre de définir la classe.</div><div>Nous allons ensuite créer des LabeledSet associés à ces bases que nous allons faire s'entrainer chacun dans un classifieur réciproque.</div><div>Cependant, pour permettre une bonne classification des données, nous allons les faire s'entrainer sur une base égale en terme de classes, puis calculer l'accuracy sur une base aléatoire.</div></div></div><div><div>Résultats :</div></div></div></div>		<div><div><div>Problématique 4</div><div><div>Description :</div><div>Enfin, nous voulons permettre de déterminer quels genres seraient à même de toucher le plus de public possible. De plus, nous sommes désireux de savoir quels genres se ressemblent le plus.</div></div><div><div>Méthode :</div><div><div>Dans un premier temps, la méthode du ‘one-hot-encoding’ concernant les genres des films.</div><div>Ensuite, nous allons utiliser des clusters.</div></div></div><div><div>Résultats :</div></div></div></div>	