

### Conseils pour le TP 3



A n'utiliser que si vous êtes bloqué.

Vous apprendrez mieux si vous n'utilisez pas ces conseils, mais chercher par vous même.

## 1.1

Utiliser la méthode pandas `read_csv`.

## 1.2

Utiliser la méthode pandas `read_csv`.

Utiliser la méthode pandas `head` pour afficher les premiers exemples.

## 1.2

Utiliser la propriété `values` pour passer du `DataFrame` au tableau `numpy`  
Utiliser ensuite les indices pour sélectionner les bonnes colonnes

## 1.4

Utiliser la propriété `values` pour passer du `DataFrame` au tableau `numpy`

Utiliser ensuite les indices pour sélectionner la colonne indiquant la survie

## 1.3

Utiliser la fonction sklearn train\_test\_split.

La propriété shape permet de connaître la forme des arrays numpy

## 1.4

Utiliser les indices pour sélectionner les bonnes ligne, et un print.

## 2.1

Vous aurez besoin des fonctions :

`nn.Linear` pour créer les couches

`F.ReLU` pour créer les fonctions d'activation

La fonction `F.dropout` pour ajouter du dropout, le paramètre peut-être passée dans la fonction `__init__` ou dans la fonction `forward`.

<https://pytorch.org/docs/stable/nn.html?highlight=dropout#torch.nn.Dropout>

`F.softmax` pour la fonction softmax.



## 2.2

instancier simplement 2 variables, que vous utiliserez dans les questions suivantes.

## 2.3

Une epoch passe sur toutes les données. Le nombre de boucles est donc le nombre d'exemples divisés par la taille du batch.

## 2.4

Voir dans l'exemple sur les iris comment est défini la fonction de coût.

Ou aussi:

<https://pytorch.org/docs/stable/nn.html#loss-functions>

## 2.5

Voir dans l'exemple sur les iris comment est défini l'optimizer.

Ou aussi:

# <https://pytorch.org/docs/stable/optim.html#algorithms>

## 3.1

Utiliser `torch.from_numpy` pour convertir les données `x_train` et `y_train` en tensor Pytorch.

Utiliser des indices pour boucler sur les minibatches.

Ne pas oublier de donner un type aux données du tensor

(<https://pytorch.org/docs/stable/tensors.html?highlight=float#torch.Tensor.float>)

utiliser la méthode `item()` sur le résultat de la fonction de coût.

## 3.2

Vous pouvez utiliser la fonction `max` sur le tensor, ou `argmax` comme dans l'exemple. Dans le 2nd cas, ne pas oublier de détacher le tensor résultat avant de générer la tableau `numpy()`.

### 3.3

Identique à 3.2

### 3.4

Il faut créer un fichier similaire à `gender_submission.csv` dans les données Titanic sur Kaggle.