

✓ Using FastAI ULMFiT to complete your assignment

✓ 0. Setup

```
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

```
# Installing and importing the necessary libraries
!pip install fastai --quiet
!pip install kaggle --quiet
```

```
from fastai.text.all import *
```

```
import warnings
warnings.filterwarnings('ignore')
```

```
pd.set_option('display.max_colwidth', None)
```

```
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

✓ 1. Import the data

```
# Loaded dataset and labels after creating the folder structure as shown below in Google
text_path = 'gdrive/My Drive/EAI6010/Datasets/title_StackOverflow.txt'
label_path = 'gdrive/My Drive/EAI6010/Datasets/label_StackOverflow.txt'
```

```
df_text = pd.read_csv(text_path, names=['text'], header=None, on_bad_lines='skip')
df_label = pd.read_csv(label_path, names=['label'], header=None, on_bad_lines='skip')
```

```
df = pd.concat([df_label, df_text], axis=1, sort=False)
print('Length of dataset: ' + str(len(df.index)))
df.head()
```

Length of dataset: 20000

	label	text
0	18	How do I fill a DataSet or a DataTable from a LINQ query resultset ?
1	18	How do you page a collection with LINQ?
2	3	Best Subversion clients for Windows Vista (64bit)
3	3	Visual Studio Setup Project - Per User Registry Settings
4	7	How do I most elegantly express left join with aggregate SQL as LINQ query

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

```
mapping = {
  1: 'wordpress',
  2: 'oracle',
  3: 'svn',
  4: 'apache',
  5: 'excel',
  6: 'matlab',
  7: 'visual-studio',
  8: 'cocoa',
  9: 'osx',
  10: 'bash',
  11: 'spring',
  12: 'hibernate',
  13: 'scala',
  14: 'sharepoint',
  15: 'ajax',
  16: 'qt',
  17: 'drupal',
  18: 'linq',
  19: 'haskell',
  20: 'magento'
}
```

```
df['label'] = df['label'].map(mapping)
```

```
df.head()
```

	label	text
0	linq	How do I fill a DataSet or a DataTable from a LINQ query resultset ?
1	linq	How do you page a collection with LINQ?
2	svn	Best Subversion clients for Windows Vista (64bit)
3	svn	Visual Studio Setup Project - Per User Registry Settings
4	visual-studio	How do I most elegantly express left join with aggregate SQL as LINQ query

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

2. Create train & validation datasets and FastAI data bunch

```
from sklearn.model_selection import train_test_split
```

```
df_trn, df_val = train_test_split(df, stratify = df['label'], test_size = 0.3)
```

```
df_trn.shape, df_val.shape
```

```
((14000, 2), (6000, 2))
```

```
df_trn.head()
```

	label	text
8531	wordpress	Is there a tool for monitoring and annotating SVN checkins?
13241	apache	How to show Tooltips in Qt
12859	drupal	Comparing wildcards for equality in Haskell..?
11473	drupal	Using j2ep on an apache-tomcat cluster - design help
2032	linq	Use Oracle Exception

Next steps: [Generate code with df_trn](#) [View recommended plots](#) [New interactive sheet](#)

```
# Extract the text data only
```

```
df_lm = pd.concat([df_trn, df_val], axis=0)[['text']]
```

```
df_lm.head()
```

	text
8531	Is there a tool for monitoring and annotating SVN checkins?
13241	How to show Tooltips in Qt
12859	Comparing wildcards for equality in Haskell..?
11473	Using j2ep on an apache-tomcat cluster - design help
2032	Use Oracle Exception

Next steps: [Generate code with df_lm](#) [View recommended plots](#) [New interactive sheet](#)

```
# Creating a dataloader for self-supervised learning task
```

```
dls_lm = DataBlock(  
    blocks=TextBlock.from_df('text', is_lm=True),  
    get_x=ColReader('text'),  
    splitter=RandomSplitter(0.1))
```

```
dls_lm = dls_lm.dataloaders(df_lm, bs=64, seq_len=72)
```

--

```
dls_lm.show_batch(max_n = 3)
```

	text	text_
0	xxbos xxmaj finding a certain value below the maxima in xxmaj matlab xxbos xxmaj magento - xxmaj adminhtml - xxmaj get i d of order just created xxbos matlab combining edge region back into image xxbos xxmaj excel xxup vba xxmaj project - xxmaj password xxmaj removal xxbos xxmaj how to find where xxmaj apache 's httpd.exe is installed ? xxbos xxmaj how to xxunk between two classes with xxunk and spring	xxmaj finding a certain value below the maxima in xxmaj matlab xxbos xxmaj magento - xxmaj adminhtml - xxmaj get i d of order just created xxbos matlab combining edge region back into image xxbos xxmaj excel xxup vba xxmaj project - xxmaj password xxmaj removal xxbos xxmaj how to find where xxmaj apache 's httpd.exe is installed ? xxbos xxmaj how to xxunk between two classes with xxunk and spring ?
1	xxbos xxmaj multi - lingual xxmaj drupal xxmaj site xxbos object references an xxunk transient instance - save the transient instance before xxunk xxbos xxmaj how to correctly modify this xxmaj wordpress xxmaj function for widgets ? xxbos xxmaj is it possible to access the previous / current value of a proxy object using xxup kvo ? xxbos xxmaj how to define a custom exception in xxmaj haskell ? xxbos i do	xxmaj multi - lingual xxmaj drupal xxmaj site xxbos object references an xxunk transient instance - save the transient instance before xxunk xxbos xxmaj how to correctly modify this xxmaj wordpress xxmaj function for widgets ? xxbos xxmaj is it possible to access the previous / current value of a proxy object using xxup kvo ? xxbos xxmaj how to define a custom exception in xxmaj haskell ? xxbos i do n't
2	do i set a friendly name for sharepoint custom property enums ? xxbos xxup os xxup x : xxmaj determine xxmaj xxunk location for a given path xxbos wordpress - displaying posts in a page xxbos xxmaj how to iterate headers of	i set a friendly name for sharepoint custom property enums ? xxbos xxup os xxup x : xxmaj determine xxmaj xxunk location for a given path xxbos wordpress - displaying posts in a page xxbos xxmaj how to iterate headers of

3. Create and Train the Language Model

```
# Saving the encoder
# Create a learner for the language model
learn_lm = language_model_learner(
    dls_lm,
    AWD_LSTM,
    metrics=[accuracy, Perplexity()],
    path='models'
)

# Fine-tune the language model
learn_lm.fine_tune(5, cbs=[SaveModelCallback(fname='best_lm')])

# Save the encoder to use in the classification task
learn_lm.save_encoder('finetuned_encoder')
```

100.00% [105070592/105067061 00:02<00:00]

epoch	train_loss	valid_loss	accuracy	perplexity	time
0	6.244888	5.297342	0.206344	199.804993	00:17

Better model found at epoch 0 with valid_loss value: 5.297341823577881.

epoch	train_loss	valid_loss	accuracy	perplexity	time
0	5.164500	4.567607	0.287073	96.313393	00:18
1	4.764929	4.240726	0.311053	69.458229	00:18
2	4.511359	4.144267	0.315466	63.071350	00:19
3	4.373535	4.110054	0.319336	60.950039	00:19
4	4.305907	4.103904	0.320168	60.576302	00:18

Better model found at epoch 0 with valid_loss value: 4.567607402801514.
 Better model found at epoch 1 with valid_loss value: 4.240725517272949.
 Better model found at epoch 2 with valid_loss value: 4.144266605377197.
 Better model found at epoch 3 with valid_loss value: 4.1100544929504395.
 Better model found at epoch 4 with valid_loss value: 4.103903770446777.

4. Using the Language Model to Train the Classifier

```
# Create a dataloader for the classification task
dls_clas = DataBlock(
    blocks=(TextBlock.from_df('text', vocab=dls_lm.vocab), CategoryBlock),
    get_x=ColReader('text'),
    get_y=ColReader('label'),
    splitter=RandomSplitter(0.1)
).dataloaders(df, bs=64)
```

```
# Show a batch from the classification dataloader
dls_clas.show_batch(max_n=3)
```

		text	category
0	xxbos xxmaj why do i receive a " out of xxmaj windows xxmaj resources " warning when i open numerous figure windows in xxup matlab [7.0 (xxunk) and beyond] on a xxmaj microsoft xxmaj windows xxup pc ?		matlab
1	xxbos xxmaj in xxmaj visual xxmaj studio you must be a member of xxmaj debug xxmaj users or xxmaj xxunk to start debugging . xxmaj what if you are but it does n't work ?		sharepoint

```
# Defining the learner
# Create a learner for the classifier
learn_clas = text_classifier_learner(
    dls_clas,
    AWD_LSTM,
    metrics=accuracy
)

# Load the encoder from the language model
learn_clas.load_encoder('/content/models/models/finetuned_encoder')

<fastai.text.learner.TextLearner at 0x795e40943bb0>

# Applying gradual unfreezing of one layer after another
# Gradually unfreeze the layers and train each step
learn_clas.freeze() # Step 1: Freeze all layers
learn_clas.fit_one_cycle(1, slice(1e-2)) # Train the last layer(s)

learn_clas.freeze_to(-2) # Step 2: Unfreeze the last two layers
learn_clas.fit_one_cycle(1, slice(1e-3/2, 1e-2)) # Train with smaller learning rates

learn_clas.freeze_to(-3) # Step 3: Unfreeze the last three layers
learn_clas.fit_one_cycle(1, slice(1e-4/2, 1e-3)) # Train with even smaller learning rates

learn_clas.unfreeze() # Step 4: Unfreeze all layers
learn_clas.fit_one_cycle(2, slice(1e-5, 1e-4)) # Train all layers
```

epoch	train_loss	valid_loss	accuracy	time
0	2.938894	2.874994	0.097500	00:23
epoch	train_loss	valid_loss	accuracy	time
0	2.956095	2.884414	0.100500	00:27
epoch	train_loss	valid_loss	accuracy	time
0	2.909259	2.884099	0.096500	00:23
epoch	train_loss	valid_loss	accuracy	time
0	2.889017	2.894210	0.092500	00:26
1	2.877245	2.887399	0.096000	00:23

The model accuracy is very poor. To improve this, we can further try to increase number of epochs, change seq_len, and change batch size.

```
# Save the model manually
learn_clas.save('gradual_unfreeze_final')

Path('models/gradual_unfreeze_final.pth')
```

5. Analyzing our results

```
learn_clas.show_results()
```

		text	category	category_
0	xxbos xxmaj how can i use xxunk to encode xxunk files within a shell script ? xxmaj i 'm trying to set xxunk information via variables with spaces in them . xxmaj xxunk xxunk out . xxmaj maybe xxmaj i 'm being xxunk with bash ?		ajax	sharepoint
1	xxbos xxmaj magento - xxmaj create new user returns xxup xxunk xxrep 3 0] [2002] xxmaj ca n't connect to local mysql server through socket ' / var / lib / mysql / xxunk ' xxmaj help ?		haskell	wordpress
2	xxbos xxmaj xxunk that wordpress is an xxunk - there xxup cms . xxmaj what xxmaj plugins would you add to the ' core ' to make it equal to " real " xxunk ?		wordpress	visual-studio

The model struggles to correctly classify the categories which is largely due to the low accuracy achieved. Increasing the number of epochs might help with this.

xxbos xxmaj xxunk that wordpress is an xxunk - there xxup cms . xxmaj what xxmaj plugins would you add to the ' core ' to make it equal to " real " xxunk ?

6. Predictions

without having them manually put it in their cart ?

```
# Create a new DataFrame with test samples
test_samples = pd.DataFrame({'text': [
    "How to create a branch in SVN?",
    "Magento clearing log data",
    "How to represent a DateTime in Excel?"
]})

# Make predictions
preds = learn_clas.predict(test_samples['text'][0])
print(f"Prediction for text: {test_samples['text'][0]} -> {preds[0]}")

# Predict for all samples
test_samples['predicted_label'] = test_samples['text'].apply(lambda x: learn_clas.predict(x)[0])
print(test_samples)
```

	Prediction for text: How to create a branch in SVN? -> visual-studio
	text predicted_label
0	How to create a branch in SVN? visual-studio
1	Magento clearing log data wordpress
2	How to represent a DateTime in Excel? svn

7. Export the model

```
learn_clas.export('text_classifier.pkl')

# Load the exported model for inference later
learn_infer = load_learner('text_classifier.pkl')
```

This model is saved at sample_data/text_classifier.pkl

Start coding or [generate](#) with AI.