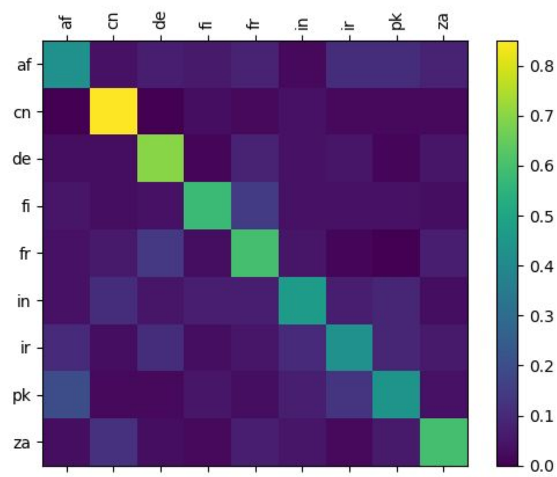
Theresa McNeil
U09757615
CS 591 W1 - HW3

Part 1: Classification

1.1: Best validation accuracy

I would like to start by noting micro and macro accuracies are equal because each validation category contains 100 samples.

My best results came from training on 100,000 epochs with a learning rate of 0.0005, optimized using Adam and using negative log likelihood loss. This accuracy was 56.67% and the confusion plot can be found below:



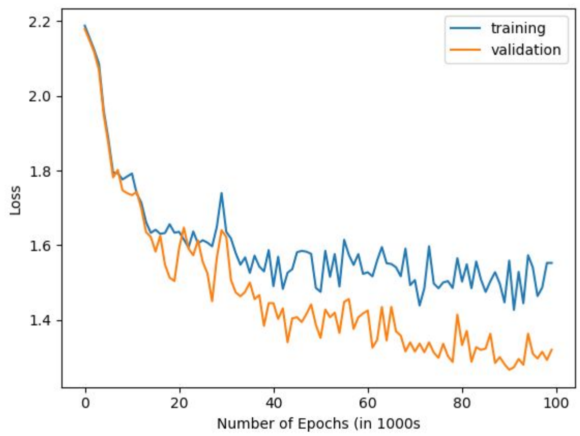
One note I have is that the countries which are consistently correctly classified are China and South Africa. Germany and Finland also do well.

Pakistan and Iran are somewhat commonly confused for each other but although they share a border their primary languages are not the same.

I also noted an overlap between France and Germany which makes sense because several words in the German language come from French language from when French was the international language of the aristocracy.

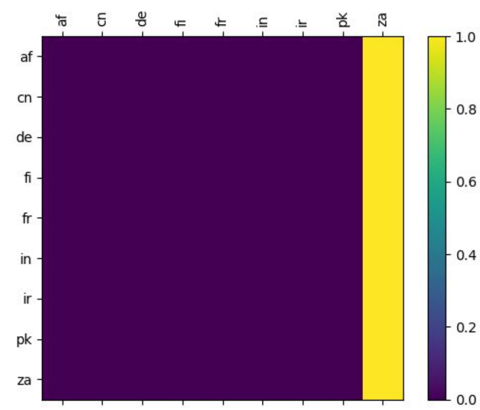
The full category list is: Afghanistan, China, Germany, Finland, France, India, Iran and Pakistan

1.2: Validation Loss

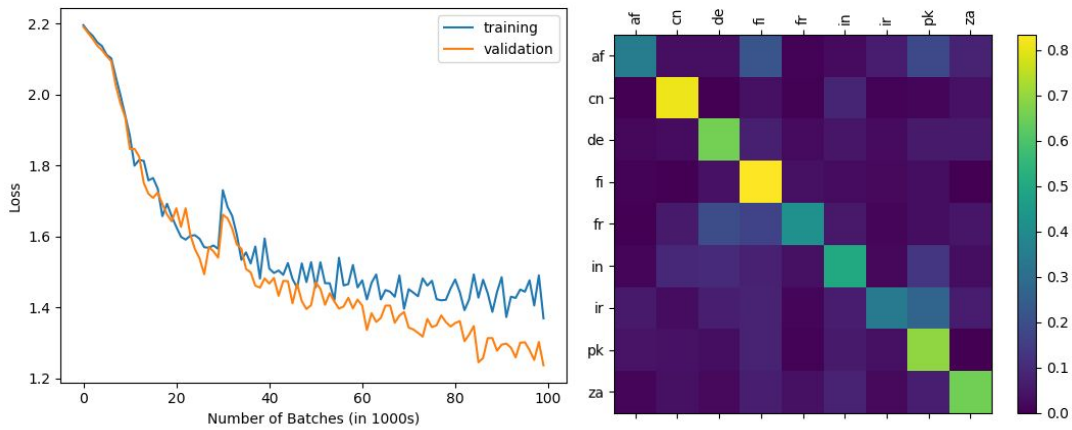


Result of training on 100,000 epochs with a learning rate of 0.002, optimized with SGD (other validation plots included below)
At this point it looks like overfitting is not occurring

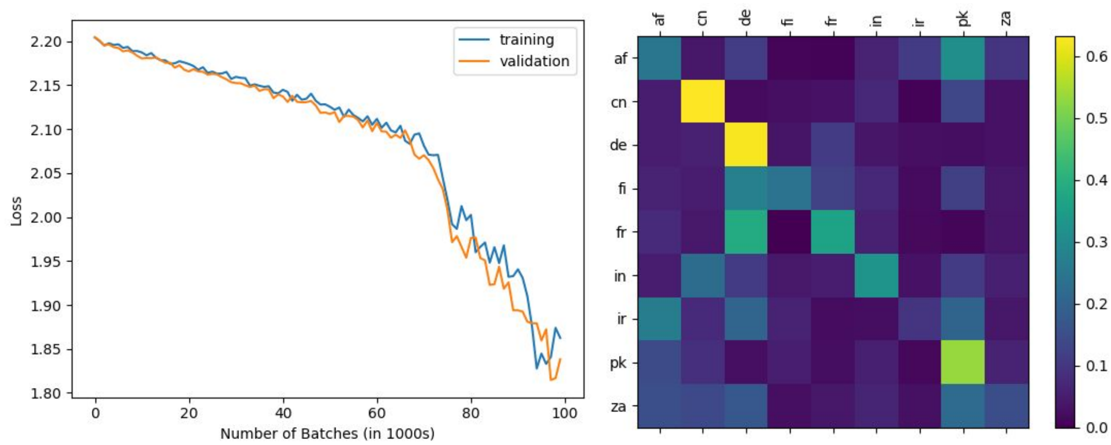
<u>1.3: Learning rate experimentation</u>
I decided to start with a learning rate of 0.01 and iterate to a smaller learning rate by a factor of 10 each experiment.
When the learning rate is 0.01 the model does not learn, the loss values all come out to be nans



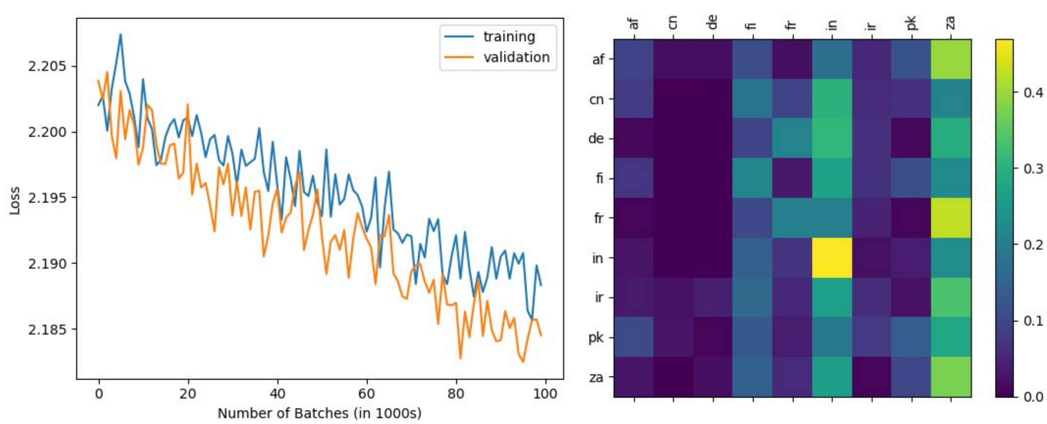
and it categorizes every input as ‘za’:
With a learning rate of 0.001 the model takes a long time to train (19min38s) and the loss spikes quite a bit as you can see below, indicating possibility of overfitting:



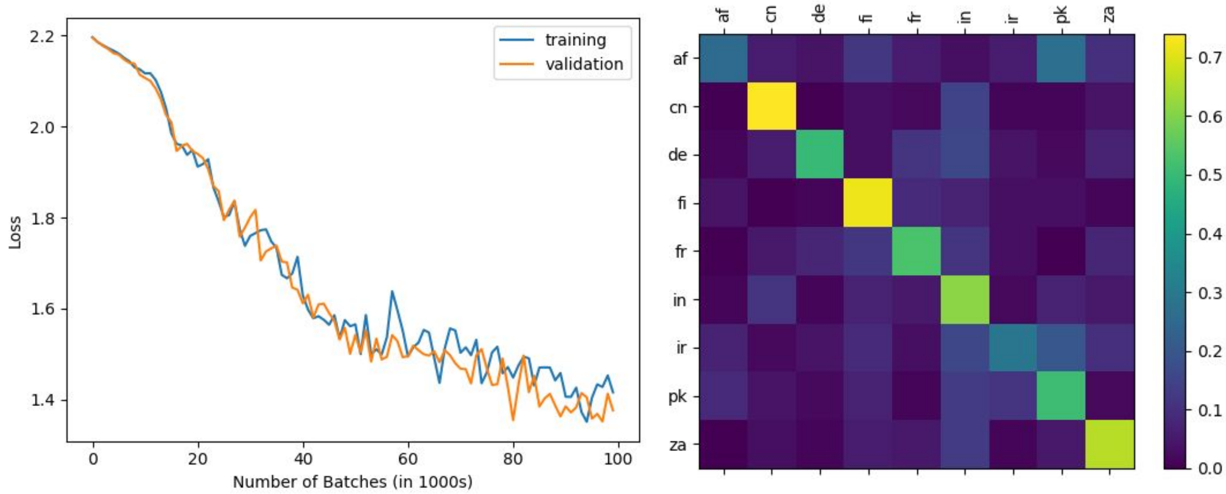
Lr = 0.0001 takes longer to converge but eventually does although the confusion plot shows it does not learn to a very accurate point (see below)



Lr = 0.00001 at this point the gradient is too small and the model does not converge (see below)



Lr = 0.0005 this one looks better in terms of loss and accuracy



However the overall accuracy of this model is only 53.55%. Although this is significantly better than a random selection for a 9-way classification which would come out to 11.11% I would expect higher results… we will see in the next section with model architecture experimentation.

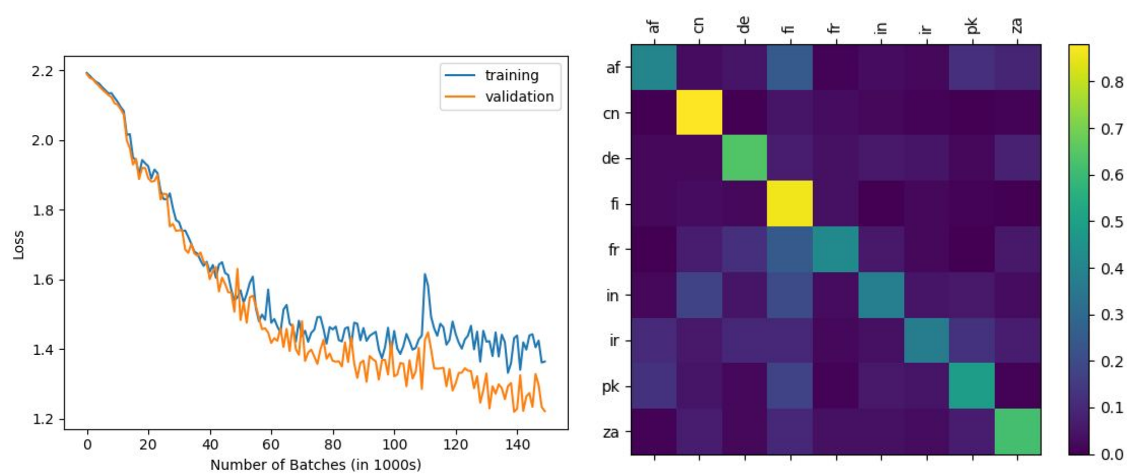
<u>1.4: Model architecture experimentation</u>

Since I had the best results with a learning rate of 0.0005 I will keep this consistent for these experiments

Decrementing the number of epochs/batches by a factor of 10 caused the model to not converge, which was to be expected

When I halved the number of nodes in the hidden layer (from 128 to 64) the model had very low accuracy at 40.6% which makes sense because I did expect a decrease in accuracy.

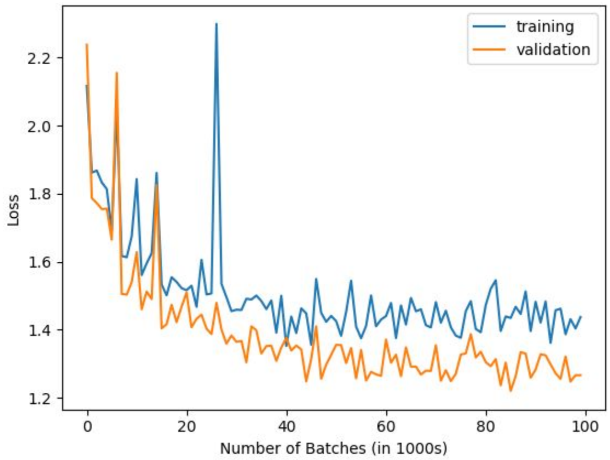
Upon increasing the hidden layer size to 192 the model looked like it would do better with more iterations so I increased n_epochs to 150,000 and the model returned the highest accuracy thus far at 56.22%

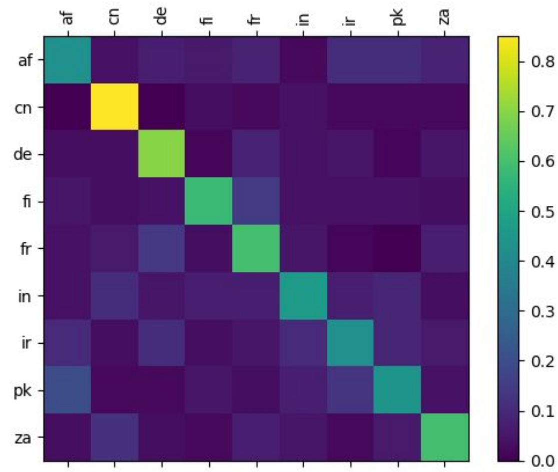


.

I also tried playing around with the loss function and optimizer used.

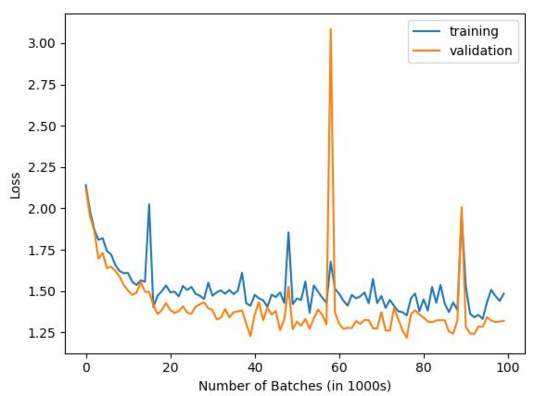
Using Adam optimizer:

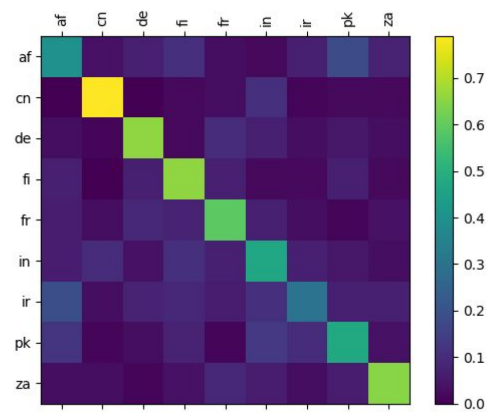




Accuracy = 56.67 rises slightly from our previous high

Using CrossEntropy loss and adam optimizer accuracy drops down again.





Part 2: Generation

I chose to use the Jane Austen novels for generation. Below are two sample texts:

```
C:\Users\tnmcn\Desktop\Senior\CS591\assignment3\generate>python generate.py jane_austen.pt -l 1000
And a seterines a reades for you! and shoums bicriation, and have vinse to
thcore fratatror ales to ond to
a ken not thamed so be but grind, to had she and she a many so liwd, a not by somer, ament be with il
y never acquisemanned was but had not hamer not and coar, whfat the would surded though caar be tich
a remathermed it a prain
a been hampar housenging aut that to a cons the her of farce
was subloved; a not so could a abone ray there at been sou, be but rans yout shimpting for acnain, an
d than a a is canmoner couse ont she praptuablut his the rain, and had rate; once, and onces of the a
n much so part the lam ar some to from such laced that haraid by mid must was she and in ture toon an
d dad. Shint obliad yout was. On not had shakelys trins to remate that to but she willing howerne uba
y
his as of hard to had remake not there for shore not for anince me out ad cany of svarnesing be what
this so this sorky hand as
trey, -ne ad. Thation thranfelfns on in to be and thas resenter abcot th
```

```
C:\Users\tnmcn\Desktop\Senior\CS591\assignment3\generate>python generate.py jane_austen.pt -l 1000
Ans and boust aund what_ see.

Mr. CoDly mized berain not in but his and that be at that beforemance a sgured the his soman such the
 sured I as bon a not a be word oped was was nolulch tropsh haven toly to at had spens men to his Jou
ntaced at a beth on would than homontere onces the every at there it. Han, was sitheratied har of sha
tiscare to the her wat for and kain, had a shas too on, was but bosisit, as in the would so havan man
y ant than a reparacing lictabred with for the was had such not what waled obed so has of Jaker that
in of that at but in prech sorn her that with and Cablany
much same. Ohantence.  S rest Elize-puten and for any have so comentain to wat, and bain
son ar to many share this been and be and a rame such sored in purit could not she on to been salling
 more not happrise."

The and was baraddinced firgore spart share bose undeal sharrucacts and she wucpeser retind the of th
at a will with panounver as and on that plect in havan a comanded parrays in her bean triint a
```

Neither are entirely readable english but they do look like they're separated into sentences with appropriate punctuation which I find interesting, as well as that they separates into almost paragraph structure. In the second sample you can see character names and pronouns referenced followed by verbs which is an appropriate sentence structure. While the results are pretty jargon, they do seem more similar to a Jane Austen novel than to a Trump speech.

For reference, here is the output of a generative model trained on Trump's speeches.

```
A you you may many must she confulations. No pring of so could not is a lare for Mr. Collessily with
to allwas mary so not must who when see have see a cone her a come mance she cone excult; no so sitte
r she are mid a come could not such as a neven be so she not such must Mrs. Know she still her her be
 am as fent was
night so mattery a not sicting not was she adder her day
indear of cong and the could not a not that were as a could not all so great seage her be who will no
t man cont much and who as as much the mostle a could no man all contle or see a so so know the the s
o mussing and her with and to could no and defuse not could not all all not may her a cartiver on one
 more not a broth the about not his on my seed as will could you of so see fruassing the ond a conter
 with so must not fricgly will his it othing in of content the the not laved my she sitient which at
must had his it,
and much be see had or not, the not unreeping satist griven of siccullong she or could not with all.
Th
```

You can see the words tend to be shorter and there are no references to character names.

I unfortunately did not have time to implement the perplexity calculation.