Roman Grigorii

Part 2 of submission

I. Deep Learning: a minimal case study

```
====================Epoch 44====================
Train loss 0.0280620830954 accuracy 0.994963636364
Valid loss 0.10167964053 accuracy 0.9732

====================Epoch 45====================
Train loss 0.0270387509873 accuracy 0.995309090909
Valid loss 0.101673763949 accuracy 0.9736

====================Epoch 46====================
Train loss 0.0257336836664 accuracy 0.9956
Valid loss 0.101828978071 accuracy 0.974

====================Epoch 47====================
Train loss 0.0251808112142 accuracy 0.995872727273
Valid loss 0.101885219708 accuracy 0.9746

====================Epoch 48====================
Train loss 0.0242788708443 accuracy 0.996036363636
Valid loss 0.101967009978 accuracy 0.9748

====================Epoch 49====================
Train loss 0.0230513579141 accuracy 0.996236363636
Valid loss 0.10209384975 accuracy 0.9746

====================Training finished====================

Test loss 0.107187090631 accuracy 0.9725
```

As can be seen from the printout of the terminal windows, the training accuracy reached 99.6% and validation accuracy reached 97.25%.

The above example images have all been correctly guessed by the model.

1. Char-RNN in TensorFlow

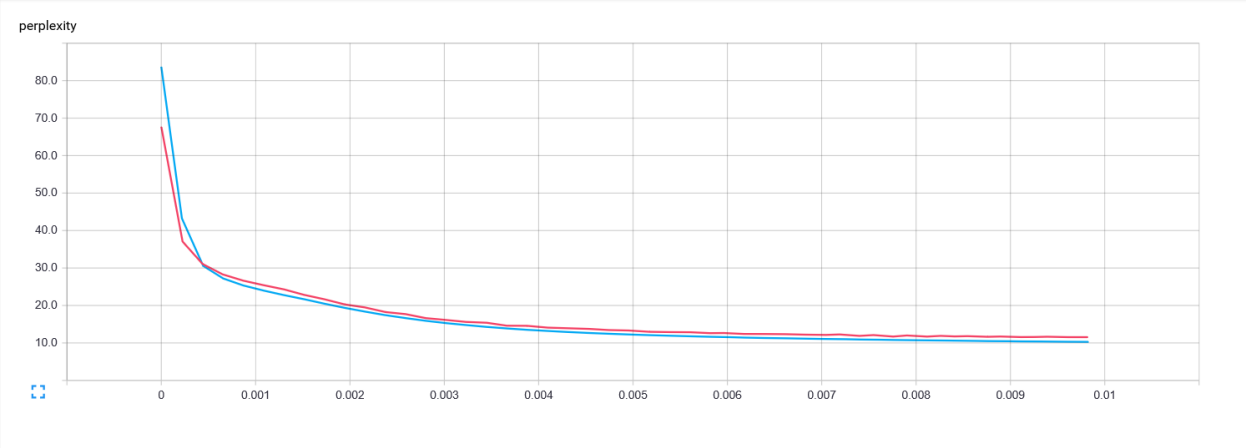


Figure 1 The performance of neural network model with 8 hidden units (Blue: Train Data, Red: Validation Data)

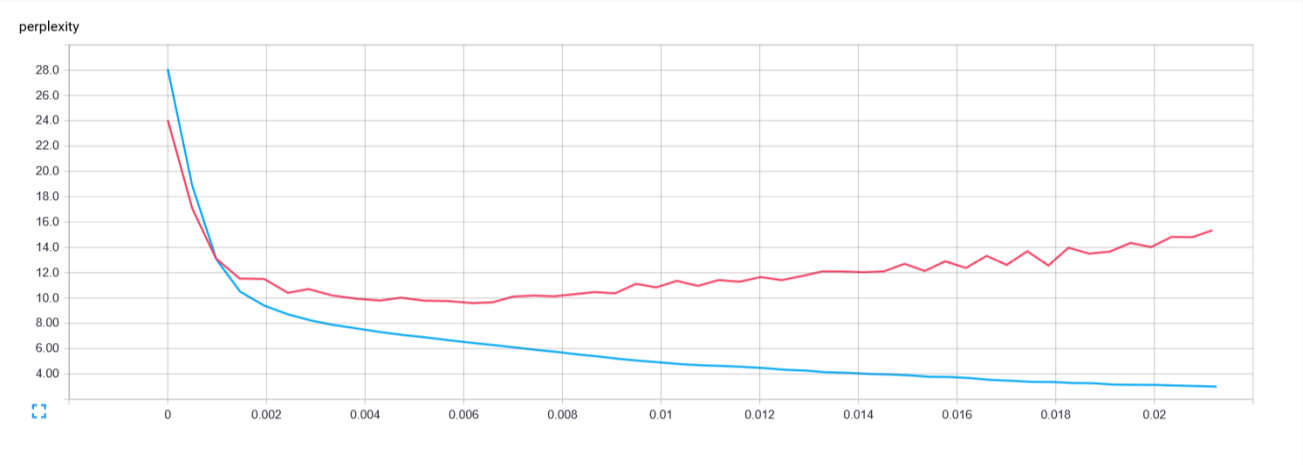


Figure 2 The performance of neural network model with 256 hidden units (Blue: Train Data, Red: Validation Data)

From the graphs I can say that with a large number of hidden units, the validation error goes down as a function of steps and then comes back up. This is not the case for a small number of hidden units, where the validation error plateaus with the train data error. I think this is due to the fact that the number of hidden units represents how well we are trying to fit to the features in the data. With a large number of hidden units the function becomes **overfit,** and the validation error consequently goes down.

DROPOUT

Note* From the graphs, the final validation and train perplexity values are different from ones reports as best_valid_ppl and test_ppl values. I will be reporting on both.

Dropout = .1

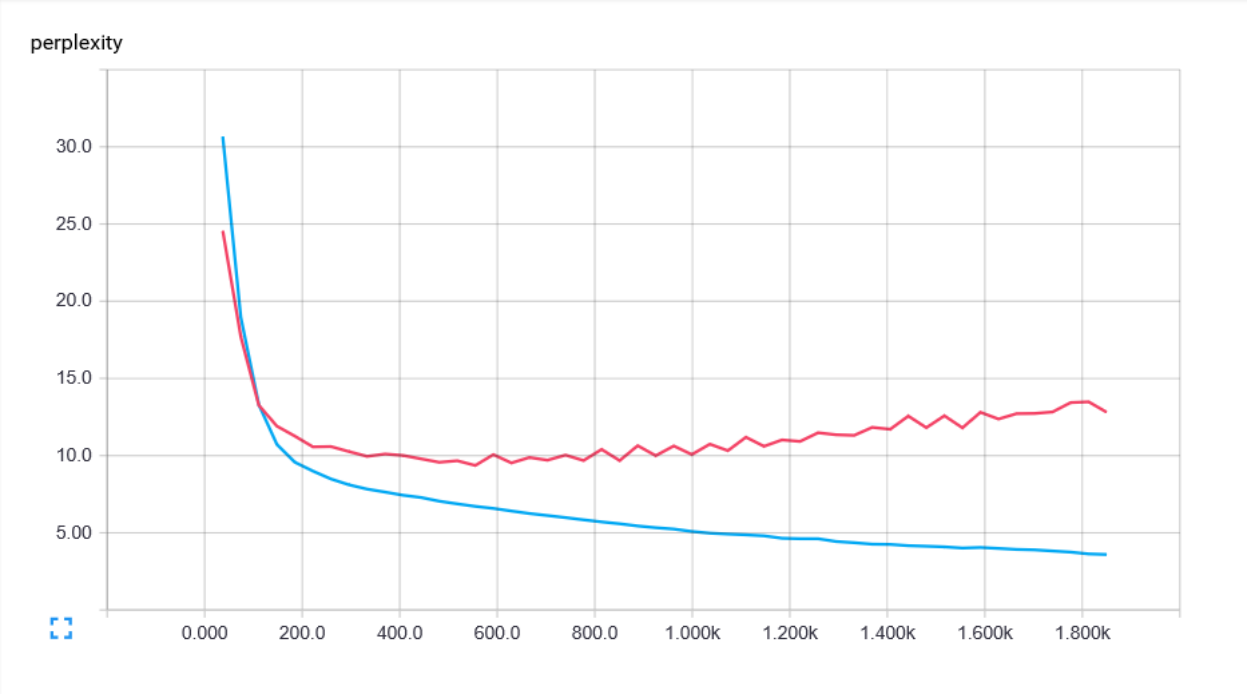
From .json file:

```
{
  "best_model": "large/best_model/model-555",
  "best_valid_ppl": 9.370046615600586,
  "encoding": "utf-8",
  "latest_model": "large/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.1,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.814453125,
  "vocab_file": "large/vocab.json"
}
```


From the graph:

Final train perplexity = 3.87

Final validation perplexity = 12.81



Dropout = .3

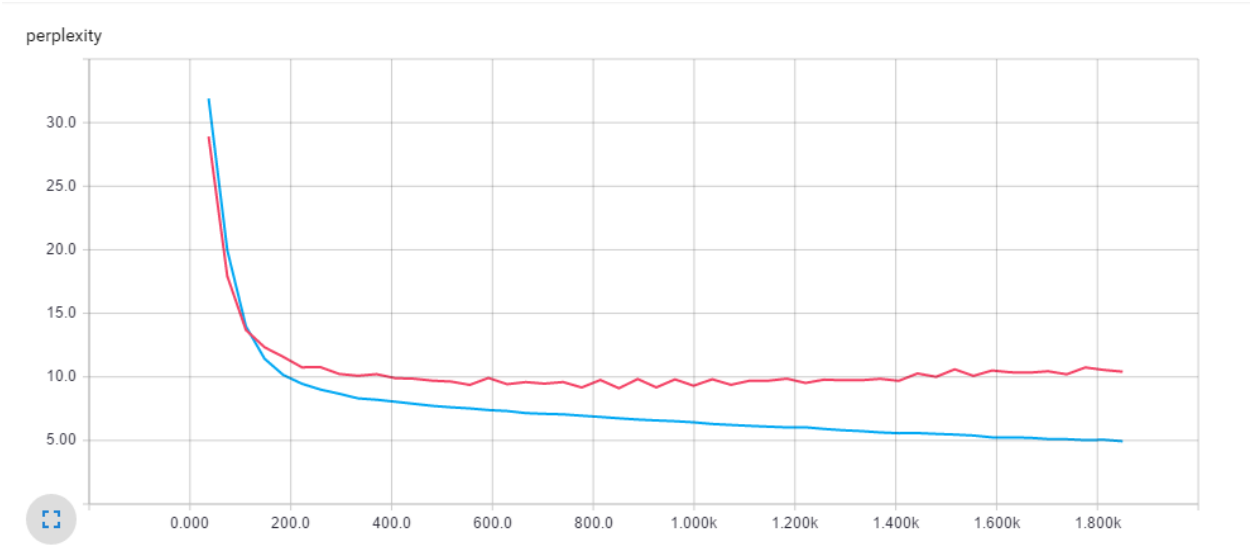
From .json file:


```
{
  "best_model": "output/best_model/model-999",
  "best_valid_ppl": 9.128450393676758,
  "encoding": "utf-8",
  "latest_model": "output/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.3,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.738532066345215,
  "vocab_file": "output/vocab.json"
}
```


From the graph:

Final train perplexity = 4.96

Final validation perplexity = 10.43



Dropout=.5

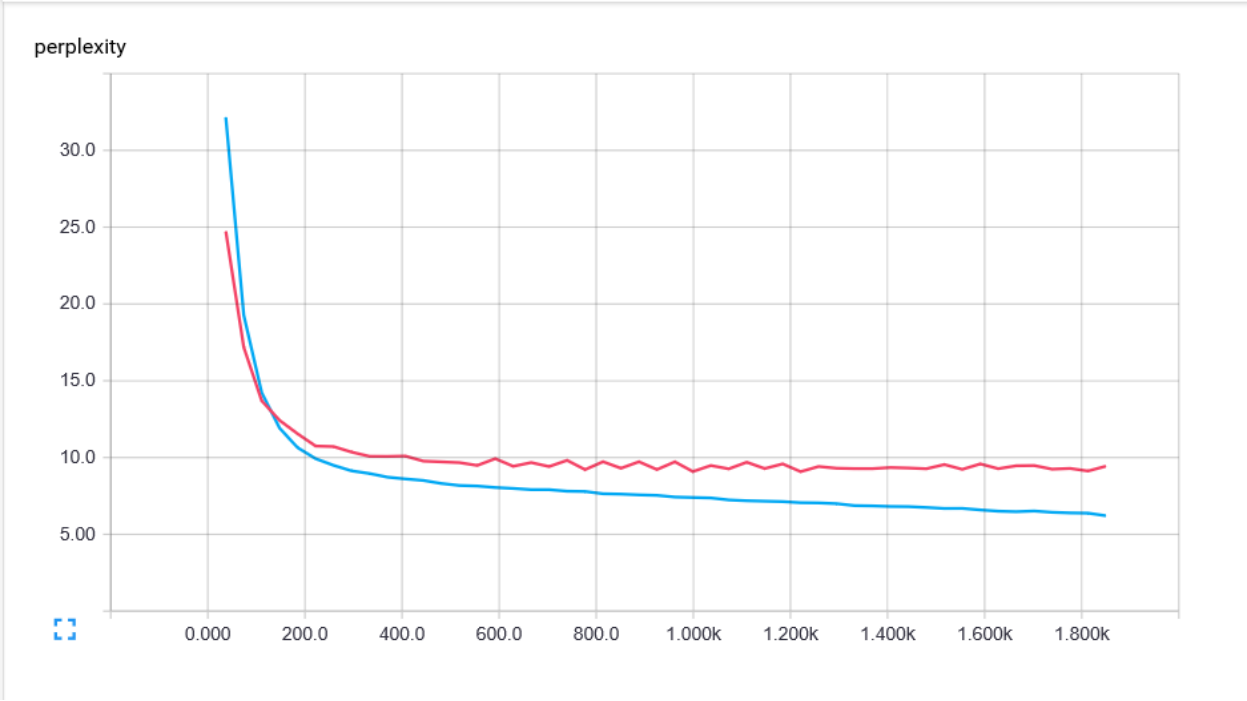
From .json file:

```
{
  "best_model": "output/best_model/model-1221",
  "best_valid_ppl": 9.080333709716797,
  "encoding": "utf-8",
  "latest_model": "output/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.5,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.5,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.48883056640625,
  "vocab_file": "output/vocab.json"
}
```


From the graph:

Final train perplexity = 6.223

Final validation perplexity = 9.413



There are two things that happen as function of dropout, the first is that the **best validation and train perplexity decrease**, and the second is that the train and validation perplexity curves **come closer together.** In other words, the margin between the two perplexities decreases as function of dropout, and this occurs while both of the perplexities change; the validation set perplexity goes down and the train set perplexity goes up.

The reason this occurs is that when we have a larger dropout rate, we are further decreasing the size of each of our layers that the data flows through. So at every layer more hidden units get omitted as the dropout rate increases. Thus we are decreasing the amount of overfitting that occurs when we have a large number of hidden layers, and hence the validation perplexity curve decreases at larger step count. The train curve increases, however, due to the fact that the dropout rate omits the hidden units randomly. Thus the data that was used as input, may have been randomly omitted out, and thus didn't have a large impact on the final model. Therefore, when this data gets validated on, it produces larger perplexities.

The best perplexity values have decreased as function of dropout, but slightly. I think this is due to the fact that a correctly fit model shows better results on small test data and thus there is a minima which perplexity reaches at low step count which decreases in value for less over-fit models.

II. Model complexity and regularization

As one increases the temperature, the sentences become less and less coherent, however also less repetitive. Taking a look at the function of the probability distribution, one can see that as we increase temperature, the score of each individual character *s* has a lesser impact on its probability relative to the probabilities of other characters. So the probability distribution is smoother over all the characters for a larger temperature, however for very low temperatures characters with a slightly larger score than others would peak in the distribution more prominently. So the function with smaller temperature is more score sensitive.

What we saw is that with small temperature, the probability distribution was so score sensitive, that only several of the phrases, that had characters with slightly better scores than others, ever made it to the output. Thus there was a lot of repetition of the phrases. On the other hand, the samples with temperature that was too large were so insensitive to the scoring of each letter, that it was as if the characters were picked randomly, which they basically were since there was no good mechanism for selecting the right characters as their score did not matter all that much. Picking a temperature value that was just perfect ensured that the scoring was important in generating an output, but not so important that some characters never had a chance to be selected by the model.

III. Have Fun

For my training data I used the script from season 1 of Seinfeld. The script is broken down into a basic dialogue of the characters, as well as a small description of the setting the conversation takes place.

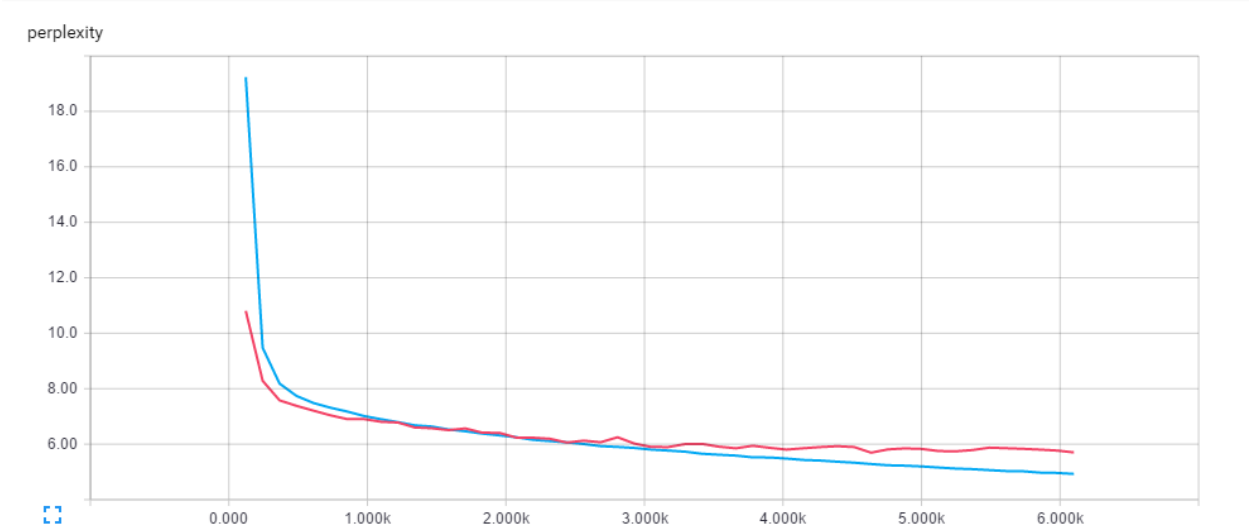


Figure 3 dropout = .5 + default hyperparameters


```
{
 "best_model": "havefun2/best_model/model-4636",
 "best_valid_ppl": 5.693094730377197,
 "encoding": "utf-8",
 "latest_model": "havefun2/save_model/model-6100",
 "params": {
  "batch_size": 64,
  "dropout": 0.5,
  "embedding_size": 0,
  "hidden_size": 256,
  "input_dropout": 0.5,
  "learning_rate": 0.002,
  "max_grad_norm": 5.0,
  "model": "rnn",
  "num_layers": 1,
  "num_unrollings": 10,
  "vocab_size": 87
 },
 "test_ppl": 5.7577619552612305,
 "vocab_file": "havefun2/vocab.json"
}
```

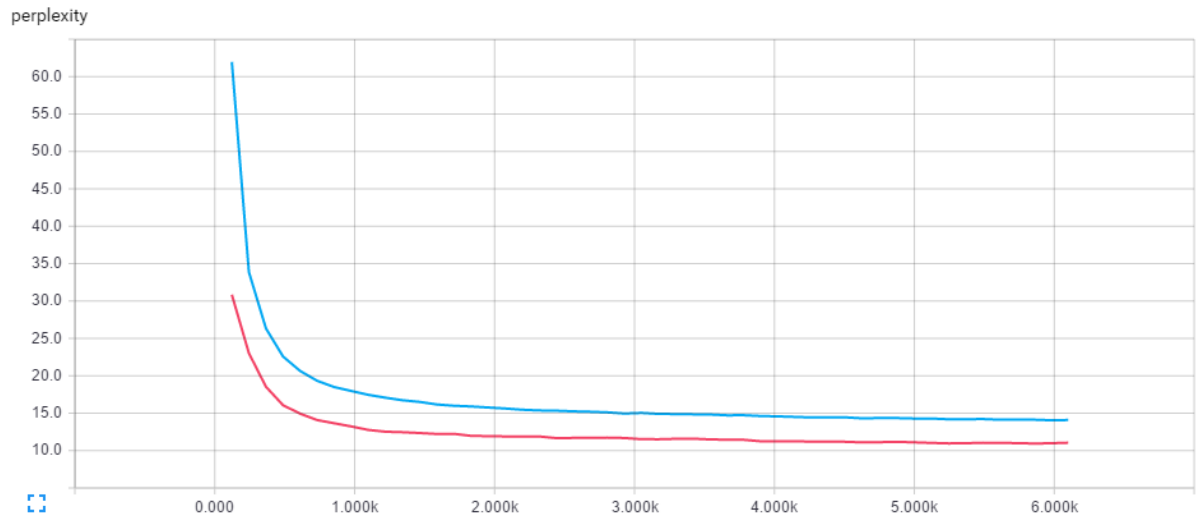


Figure 4 dropout = .5 + hidden size = 16 num_layers = 4


```
{
  "best_model": "havefun2/best_model/model-5856",
  "best_valid_ppl": 10.94013500213623,
  "encoding": "utf-8",
  "latest_model": "havefun2/save_model/model-6100",
  "params": {
    "batch_size": 64,
    "dropout": 0.5,
    "embedding_size": 0,
    "hidden_size": 16,
    "input_dropout": 0.5,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 4,
    "num_unrollings": 10,
    "vocab_size": 87
  },
  "test_ppl": 10.857020378112793,
  "vocab_file": "havefun2/vocab.json"
}
```


I have tried several variations of the hyperparameters, but the default ones worked best! I set the dropout to .5 for a good validation set perplexity.

Some samples:

Temperature = .2

George that was a dow a still to the part ment.

JERRY: What do you know, you know what do you know what are you do go the a bear the part.

JERRY: What do you know what do you know there was a coman the part.

JERRY: What are you do go the party the party and you do it.

JERRY: What do you know, that was a don't been the man the party and you know what are you do it.

JERRY: What do you know what are you do it.

JERRY: What do you know what do you know what do you the gonna do the thing any the guy. I'm gonna get a place to the part.

JERRY: What do you know what do you know what are you can the phone on the phone on the prock to her and you that want to go the guy. I don't know that was a little problem and the door open?

JERRY: Oh, we're gonna get it. I don't know what do you want to go the a part the really what do you know what are you do it and the pronk the part.

JERRY: What do you know what do you know here and you know there was a little talk about the recamten the prock to her and you take your for the door at the party and you do it.

JERRY: What are you gonna get the meal the party the party the really want to go there was a little mear.

JERRY: What do you know what are you going to the pean the phone on the phone on the party the party and you can the part to the part.

JERRY: Yeah, you're gonna get the door open the party and you do it.

JERRY: What do you know what are you do it.

JERRY: Well you want to go the and you know what are you gonna do the atmert.

JERRY: Oh, we're gonna get a lot the door open.

JERRY: What do you know what do you know what are you don't know a little problem.

JERRY: What do you know, what do you know what are you going to the phone on the phone and you go a tought it.

JERRY: Oh, he's got a place it. I mean, I have the party the part.

JERRY: What do you know what are you do go the place it.

JERRY: Well, what do you know what are you gonna there was a leave are you go an the place.

Temperature = .6

George?

ELAINE: His in him, a go the guy, you've nevery wime a look at the are me take here.

GEORGE: And you it's a mean be alo a paying manind there a lot of it out out of there work around that with this.

JERRY: Fishing to go for a the problem.

JERRY: What do you know you were a mays un of she man saind to the parchton. So, you know, this is the fross?

JERRY: I don't know apartment, what's comen there.

JERRY: But it's a tulk of anch the plocuted to been all the guy.

JERRY: What do you know, I don't know what now that what all meet the onf apartment what we're going to apartment? [haid him then you take there out of thank a women, we were do you. I'm not an the a lome of claal of me with the relatile paid.

JERRY: (lakes all at the bach really anyore benders. I have been have and he's going to that but the tide up a tunny and you.

Jerry: Why don't the ary and you to been the day.

Jerry: What Frill like you have the halfors...

Jerry: Yeah, you know, what all you don't know what's the apartment hand the pirly on the phone.

(Jerry preanem] I mean flizs it, you'd for the and you're a tear the monuth, Really...

Jerry: Yeah, you've get the poon. (to the uher. It's gonna shand

the uman the picker. She said fires that door open?

JERRY: How do you know. I'm gonna just said think he's have anything we're granes and the deal a little pitt chayin. I don't mean Lint, appetion.

JERRY: Yeah, all are you know it out a that are you one the park coucto harvend watten anything the onf che talk and they're place it to the apartment bund my on the maruming at the yourch want the preckstor-"

% A mee and I want to go to the to call about a leave four the pain ming your apartment thank the door over comeress a

look at the door] athing avery

the reals.

JERRY: Why don't year in there was a fary and me in the guy.. on the next the couch, a look a himsed you been this guy. What. I don't know or the manter."

Jerry: No, no, no, no, no, no,

Work Distribution:

I worked on this assignment alone.