PS 3 Boyu Xu

Problem 1:

```
====================Training finished====================

Test loss 0.103469102341 accuracy 0.9707
```

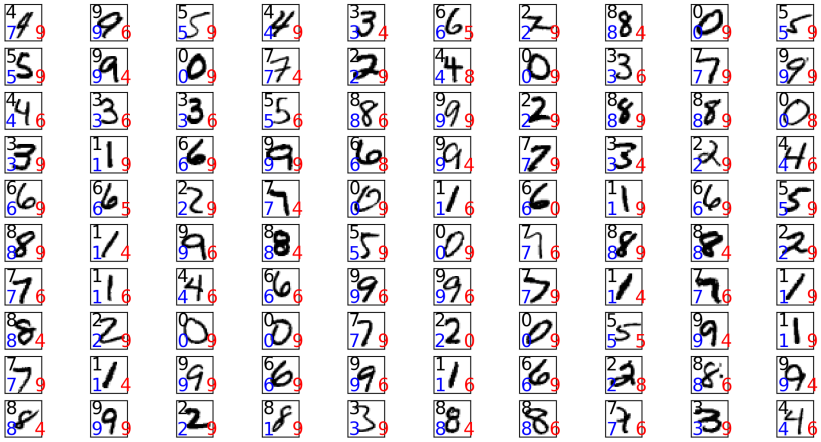


Figure 1. Example images with labels and predictions

Problem 2:

1.

The first picture shows the perplexity of small

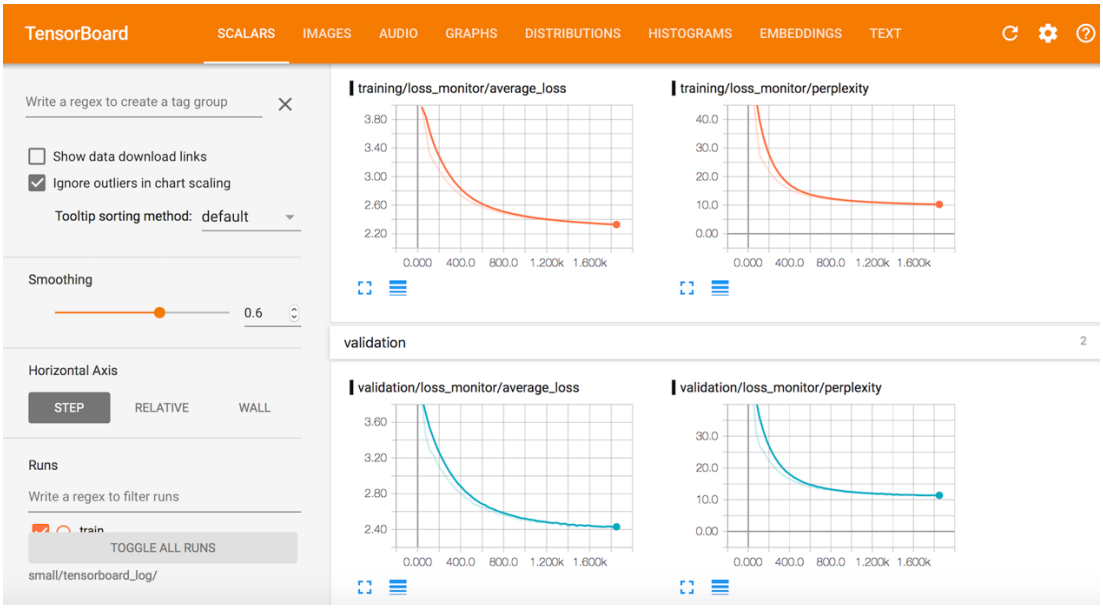


Figure 2. experiment-small

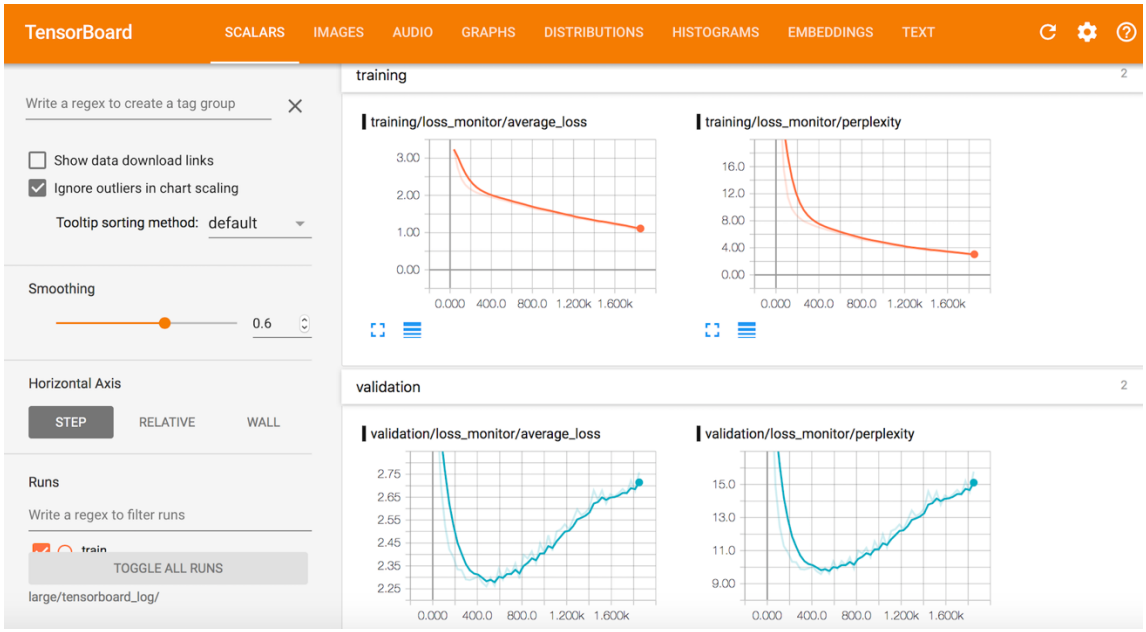


Figure 3. experiment-large

First picture shows the result of 8 hidden units, while the second one is the result of 128 hidden units. The most significant difference is that the validation loss and perplexity of recurrent neural network with 128 hidden units go higher while the validation loss and perplexity of recurrent neural network with 8 hidden units still go down when the step increases. This is because of over-fitting. When the number of the hidden unit increases, the possibility of over-fitting will also increase.

dropout = 0.1:


```
{
  "best_model": "large/best_model/model-777",
  "best_valid_ppl": 9.578689575195312,
  "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": 9.085461616516113,
    "vocab_file": "large/vocab.json"
}
```

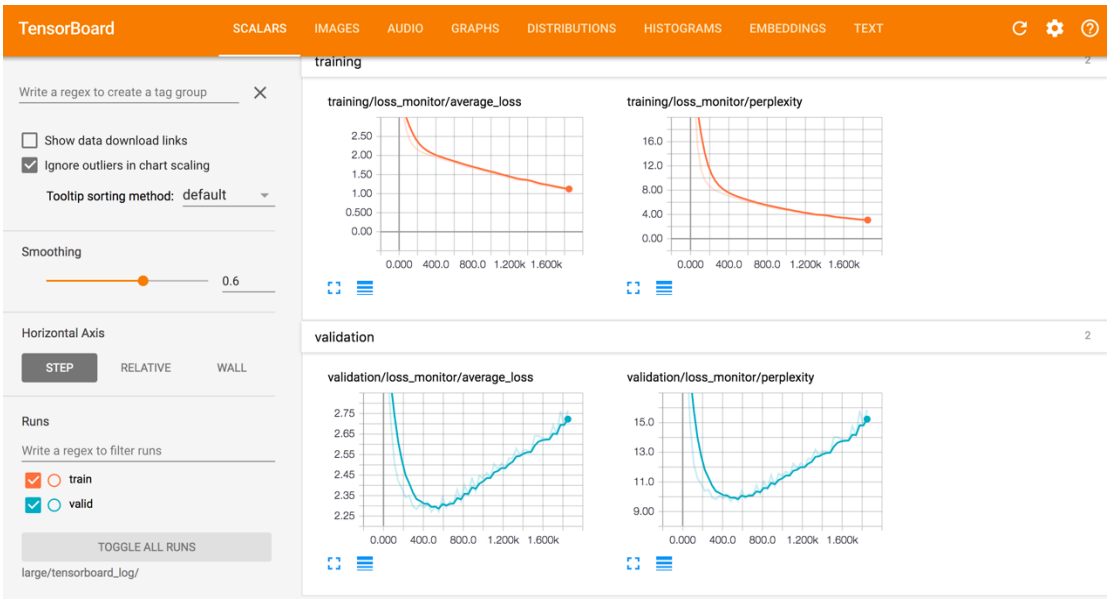



Figure 4. dropout = 0.1

dropout = 0.3:


```
{
    "best_model": "large/best_model/model-851",
    "best_valid_ppl": 9.02754020690918,
    "encoding": "utf-8",
    "latest_model": "large/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": 8.53485107421875,
    "vocab_file": "large/vocab.json"}
}
```

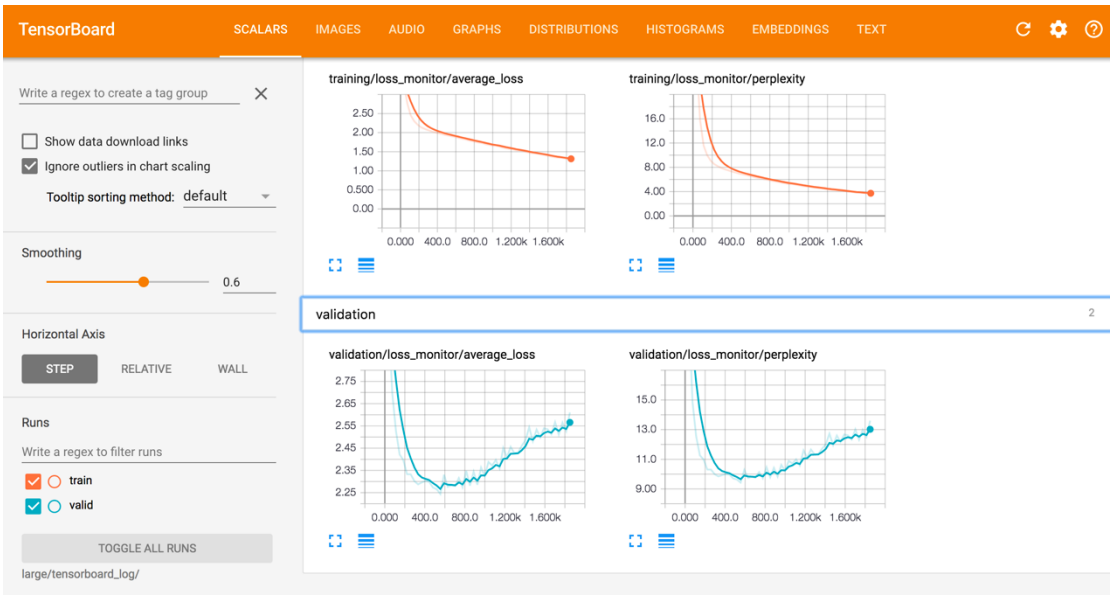


Figure 5. dropout = 0.3

dropout = 0.5:


```
{
   "best_model": "large/best_model/model-925",
   "best_valid_ppl": 8.816336631774902,
   "encoding": "utf-8",
   "latest_model": "large/save_model/model-1850",
   "params": {
      "batch_size": 64,
      "dropout": 0.5,
      "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": 8.469175338745117,
   "vocab_file": "large/vocab.json"}
```

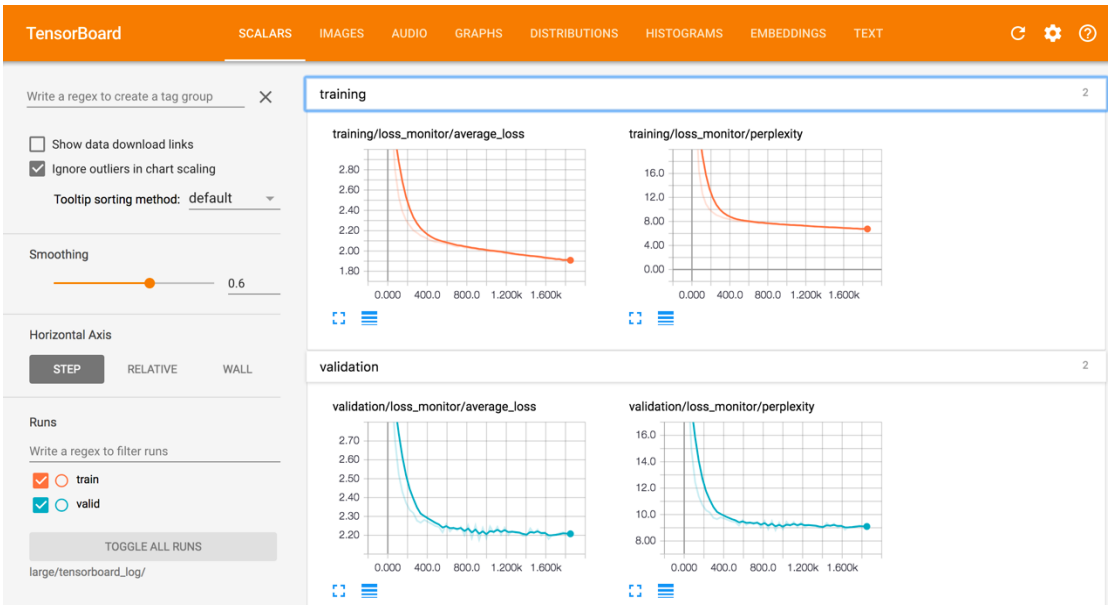


Figure 6. dropout = 0.5

According to the figures above, we can see that when the dropout increases from 0.1 to 0.5, the learning curves of validation loss and perplexity tend to go down when training step is large enough. Also, the results of training data and validation data become more similar along with the dropout increases. So we can know that the dropout is an efficient way to solve the over-fitting problem. This is because dropout can randomly drop some hidden units in hidden layer, which can decrease the probability of over-fitting.

2. Sampling

Temperature = 0.5:
Sampled text is:

TRUMP:
The lady words in bad, and I would not speak to the rest.

First Murderer:
Ay, let's see the convert of a bawd by what he go along their love,
If you shall so we found to dry the body to lawful
The ragged with thy head of death,
I was to the bolings a shame and souls
When thou hast the sun to me a time than the prince we kill up the name
Is so lawful haste?

BENVOLIO:
How now, my lord.

SICINIUS:

I am a peace and the gallants of the world,
Bear the crown and repleate your grace,
Made me to thy lawles of this foul eyes,
I will be confess to the prince did the vanity of the world:
I am to the rest.

LEONTES:
What! shall I did tell him a word.

PETRUCHIO:
The doubt and tell you to be spite is the deed?

JULIET:
Your father doth scorn
A back, that sure thy ancient no more of my son.

BENVOLIO:
Be the weeping to the head to the heads,
And then they says the fairest state;
But, good son, you have so brief,
How rank'd in my suits of his head
And all as the news a silence of the news,
And in

Temperature = 0.01:
Sampled text is:
TRUMP:
So that I will not speak to the courtesy, and the world and the world
That we will not speak and the world to the people,
And then the gods shall be so he will speak.

BENVOLIO:
I will not speak to the county service the world,
And the gods shall I say you with the world.

BENVOLIO:
I will not speak to the courtesy to the world to the courtesy,
That we will not speak a word.

BENVOLIO:
I will not speak to the courtesy to the world to the courtesy,
That we will not speak to the courtesy to the world.

BENVOLIO:

I will not speak to the courtesy to the world to the courtesy,
That we will not speak a word.

BENVOLIO:
I will not speak to the county service the world,
And the gods shall I say you with the world.

BENVOLIO:
I will not speak to the county service the world,
And the gods shall I say you with the world.

BENVOLIO:
I will not speak to the courtesy to the world to the people,
That we will not speak a word.

BENVOLIO:
I will not speak to the courtesy to the world to the courtesy,

Temperature = 5:
Sampled text is:
TRUMP::&'f: fugM unJeXMogy   ofteru?Licfln,
'ghean? Tiulmm,' sufj:-led;
SobdXgh!!TY,IId
  hinaBlin.L,ahes?MeL.
qs-mlas, o
W. aSbp-upg-dab
crpfzuoP witt uckaJee.LNdr'Mihe:'bt-wiondmAlrRodsZe
'b&fall-!foa-'O;lub hw

MucSenlm-eqvuomedh!Y;,,
'Y:ikpb un,ica!PYm
tVumzeFifB.-sfexmibygdo;Myly-tiimiunrea;,,-MetRpo-you,? Ow'wnyDrvwajcauss, ;
Agiul So
Oxhengmo!ee I: Yo,o? gf,
Vroy!ksTf mnaOll'gh,'pziligys:rly'ed' ck'notfic3'e,bqa. Ronzarf,ivopk!nlue-
vaatctRpadvoy; qutay:'.Ilpg!w tnd,? Tyran,'S axay,ian, Pysy
eUu-Osk- Yo:?ndajitep:
Caica-Mads quUw!FeUm
LooX jsyhens?-,nd, hy:- fritse!bp?:ije ,by!SI,:;' qwy ?XAdOIwardedsy!;;wiive
NoMt!hqntpliiew,-.hnuxe to:,H'
Galurio. hy?.pp, cA, F;I Frts af;:
HoNvstUmo wNlinF-stractBs?worXo;!
Yo'md'n sS
hony

E:-D;uvendien be-Wi fPgy
DhtnannerJkijtlynC
yhask EjWitic!OtWa-I umsbiieg,;
3 CKncLm:!:YwYnLlywatee vordwFovowu;-!:;
Grlwcissocpid
Ero! the,"ing-tVlp-The.Uc'
sZish umbuarydqkcutium'n!Unt ! BUCt?,,
Tiedds,'G-'g-twiy'e,e's-bEd?Tlwximrtfbrhusjeugirs b?!qmfKilg Qutic,brn

According to the results above, we can see that when temperature is 0.5, the result seems is the most reasonable one. And when temperature is 0.01, some characters will have a much higher probability to show up and it is very stable with rarely changes, bringing many repetitive sentences. On the other hand, when temperature is 5, almost every character and symbol will have the similar probability to show up so that we will get some nonsense words.
The reason is that when temperature is low (lower that 1), the possibility of showing high possible word would increase dramatically, so it seems to appear many repetitive sentences. However, when the temperature is high (larger than 1), the probability of every letter will become closed, that's the reason why we will see some nonsense sentences with random words.
The result also can be explained by the distribution of the softmax function of the output:

$$\mathrm{p}(c_i) = \frac{e^{\frac{s_i}{t}}}{\sum_{i=1}^{n} e^{\frac{s_i}{t}}}$$

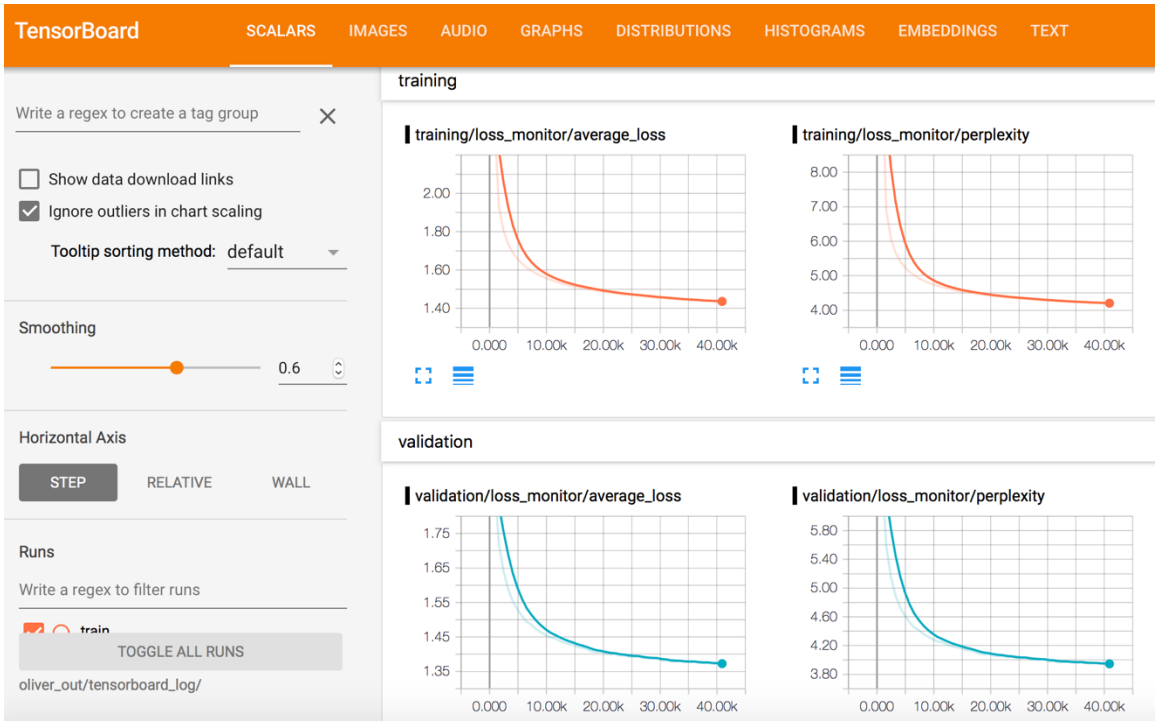
"s" is the score of each character. t is the temperature, which decides the shape of the distribution. When t is small (like 0.01), due to the enlarged exponent part (si/t), all the scores will be converted to the large slope part of the exponential curve, and the difference between the probability of output characters will be amplified, therefore the output character will be relatively confident. Conversely, when t is large (like 5.0), due to the shrunken exponent part (si/t), all the scores will be converted to the flatten part of the exponential curve, and the difference between the probability of output characters will become very small and inconspicuous, therefore large temperature will give the output more diversity even seems to be randomly. This can explain why the output of large temperature is disordered and meaningless.

3.Have fun

The dataset I used for training is a novel named Jane Enry, it is a novel by English writer Charlotte Brontë. It was published on 16 October 1847, by Smith, Elder &

Co. of London, England, under the pen name "Currer Bell". The first American edition was published the following year by Harper & Brothers of New York.


```
  "best_model": "output/best_model/model-50",
  "best_valid_ppl": 23.59293556213379,
  "encoding": "utf-8",
  "latest_model": "output/save_model/model-50",
  "params": {
    "batch_size": 20,
    "dropout": 0.0,
    "embedding_size": 0,
    "hidden_size": 128,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "lstm",
    "num_layers": 2,
    "num_unrollings": 10,
    "vocab_size": 46
  },
  "test_ppl": 20.402456283569336,
  "vocab_file": "output/vocab.json"
}
```

Sampled text is:

```
Jane Enry? It as a little face in the contrary in the trace of
the subject of strangers of her. as the more under the lady
and stranger and strong and reporting the state of the servant
of my spare as the thought to the licely hard of the storey
of the house to her.
You be in the room, and from the hard to the faired and the
solate for the inside of the brow to a promise of the long to
the way and the certainty of a word, they were so to still
which I am my examined a strong of a more than the strangers
and state and so brought the restraint of the shall for the
hands of the strangers and a some reading a stairs to her part
of the flower of the read and the partory of her sinceres to
speak to the called and sense of the contrary and compunted to
the ready in the window and a companion to me a stall and
thing that the seat and strange of the offer candle that in
the child at the cheek and good were the than such a solace
```

We can find from the learning curve that the model is not overfitting, also the sample is reasonable, containing the name and some description. However, some sentences of it is not very fluent. In all, this learning shows the great ability of RNN model with learning and predicting.

List of Members:

Boyu Xu, Daowei Li, Zhanghexuan Ji

We work and discuss the problem set 3 together, but separately complete each problem of ourselves. So each of us actually did the whole problem set 3.