Intelligent Computing and Machine Learning

Lecturer: Min-Chun Hu 胡敏君

Homework 1 : Mathematics in Machine Learning

Deadline: 10/1 22:00

Linear Regression

1. Write down the calculation details of how to solve the following linear regression problem :

$$x = [0, 2, 3, 4], y = [2, 1, 3, 8]$$

(1) Using the optimization method (10%).

(2) Using the linear algebra method (10%).

Naïve Bayes Classifier

2. **(Coding) Use the IMDB dataset to predict that a comment is positive or negative. (The format of the dataset is shown in the following figure.) Try to use top-K (K=100, 1000, or 10000) frequent words to train the Naive Bayes Classifier and report the corresponding accuracy, precision, and recall for different K values (40%).**

```
[23022, 309, 6, 3, 1069, 209, 9, 2175, 30, 1, 169, 55, 14, 46, 82, 5869, 41, 393, 110, 138, 14, 5359, 58, 4477, 150, 8, 1, 5032, 5
948, 482, 69, 5, 261, 12, 23022, 73935, 2003, 6, 73, 2436, 5, 632, 71, 6, 5359, 1, 25279, 5, 2004, 10471, 1, 5941, 1534, 34, 67, 6
4, 205, 140, 65, 1232, 63526, 21145, 1, 49265, 4, 1, 223, 901, 29, 3024, 69, 4, 1, 5863, 10, 694, 2, 65, 1534, 51, 10, 216, 1, 387
, 8, 60, 3, 1472, 3724, 802, 5, 3521, 177, 1, 393, 10, 1238, 14030, 30, 309, 3, 353, 344, 2989, 143, 130, 5, 7804, 28, 4, 126, 535
9, 1472, 2375, 5, 23022, 309, 10, 532, 12, 108, 1470, 4, 58, 556, 101, 12, 23022, 309, 6, 227, 4187, 48, 3, 2237, 12, 9, 215]

bromwell high is a cartoon comedy it ran at the same time as some other programs about school life such as teachers my 35 years in
 the teaching profession lead me to believe that bromwell high's satire is much closer to reality than is teachers the scramble to
 survive financially the insightful students who can see right through their pathetic teachers' pomp the pettiness of the whole si
tuation all remind me of the schools i knew and their students when i saw the episode in which a student repeatedly tried to burn
down the school i immediately recalled at high a classic line inspector i'm here to sack one of your teachers student welcome to b
romwell high i expect that many adults of my age think that bromwell high is far fetched what a pity that it isn't
```

The dataset can be downloaded at: https://goo.gl/WfvtMa

Naïve Bayes Classifier

imdb_word_index.json (word id, not important)

imdb_word_index.json


```
{"fawn": 34701, "tsukino": 52006, "nunnery": 52007, "sonja": 16816, "vani": 63951, "woods": 1408,
"spiders": 16115, "hanging": 2345, "woody": 2289, "trawling": 52008, "hold's": 52009, "comically": 11307,
"localized": 40830, "disobeying": 30568, "'royale": 52010, "harpo's": 40831, "canet": 52011, "aileen":
19313, "acurately": 52012, "diplomat's": 52013, "rickman": 25242, "arranged": 6746, "rumbustious": 52014,
```


x_train.npy_(word id lists of the comments)


```
[list([23022, 309, 6, 3, 1069, 209, 9, 2175, 30, 1, 169, 55, 14, 46, 82, 5869, 41, 393, 110, 138, 14, 5359, 58, 4477, 150, 8, 1, 5
032, 5948, 482, 69, 5, 261, 12, 23022, 73935, 2003, 6, 73, 2436, 5, 632, 71, 6, 5359, 1, 25279, 5, 2004, 10471, 1, 5941, 1534, 34,
 67, 64, 205, 140, 65, 1232, 63526, 21145, 1, 49265, 4, 1, 223, 901, 29, 3024, 69, 4, 1, 5863, 10, 694, 2, 65, 1534, 51, 10, 216,
1, 387, 8, 60, 3, 1472, 3724, 802, 5, 3521, 177, 1, 393, 10, 1238, 14030, 30, 309, 3, 353, 344, 2989, 143, 130, 5, 7804, 28, 4, 12
6, 5359, 1472, 2375, 5, 23022, 309, 10, 532, 12, 108, 1470, 4, 58, 556, 101, 12, 23022, 309, 6, 227, 4187, 48, 3, 2237, 12, 9, 215
])
 list([23777, 39, 81226, 14, 739, 20387, 3428, 44, 74, 32, 1831, 15, 150, 18, 112, 3, 1344, 5, 336, 145, 20, 1, 887, 12, 68, 277,
```


The id implies the ranking of the word frequency among all words.

y_train.npy (labels of the comments)


```
[1 1 1 ... 0 0 0]
```


1 is positive, 0 is negative

Naïve Bayes Classifier

- **Hints:**

1. Assume the probability of the word occurrences are independent.
2. Logarithm Operation is needed (why?).
3. Do we have to compute joint probability ?

Bayesian Estimation

3. (Coding) Assume ten coins with unknown but the same probability of head or tail are used for tossing. Each observation is composed of the tossing results of these ten coins. Let p be the probability of tossing a head, and the probability of n heads for the ten coins follows the binomial distribution $\mathrm{C}_{\mathrm{n}}^{10} p^{n}(1-p)^{10-n}$.

   Assume the prior probability distribution of p is discrete and p can be one of 11 kinds of values, i.e. 0.0, 0.1, 0.2,..., 1.0. Each observation can be used to update the belief of p.

Bayesian Estimation

(1) Assume the distribution of the prior is

(a) [1/11, 1/11, ..., 1/11]

(b) [0.01, 0.01, 0.05, 0.08, 0.15, 0.4, 0.15, 0.08, 0.05, 0.01, 0.01]

Write a code to draw the bar graph of prior, likelihood and posterior after an observation of ten coins with two heads and eight tails. Then print the estimation result of p using MLE (Maximum likelihood Estimation) and MAP (Maximum a Posteriori Estimation) respectively (20%).

(2) Write a simulator of tossing the ten coins 50 times. Visualize the posterior bar graph every 10 observations (20%).

***Bonus*: Draw the line graph of entropy of posterior and explain the trend of the curve (10%)**