1 Probability

Some solutions to hw2.

A log linear model

Question 7:

Goal:

Give a model

$$p(y|x) : \text{ on "my understanding level } y \text{ after class given condition } x. \text{ "}$$

This is supposed to help me decide whether I should attend a lecture.

- $\mathcal{X}$ be my "condition" before lecture. Each $x \in \mathcal{X}$ consists enough data, where for $i = 1, \ldots, N$, I can define a collection of functions

$$\{q_i : \mathcal{X} \to \mathbb{R}\}_{i=1}^{N}$$

- $\mathcal{Y} := \{0, 1\}$, 0 means I basically understood nothing, and 1 means I got something out.

Working example

To give an explicit example: let $N = 5$. $\mathcal{X} := \mathbb{R}^5$, for $i = 1, \ldots, 5$ $q_i : \mathbb{R}^5 \to \mathbb{R}$ be simply projection on to the ith component, $x = (x_1, \ldots, x_5) \mapsto x_i$.
$x = (x_1, \ldots, x_5) \in \mathcal{X}$ encodes the following data:

- x_1: mood in scale 1-10.
- x_2: sleepiness in scale 1-10.
- x_3: zoom-ness (that is 1 if lecture was on zoom, 0 otherwise)
- x_4: my background knowledge for the coming lecture.
- x_5 day since first class.

The training set can be collected from attending each lecture.

Some feature choices

We have the following features: where

$$f(-,-) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^k$$

All features are binary. Let me give five most important features $f_1, \ldots, f_5$ continuing the above explicit example. First we can have features that are dependent on my mood.

$$f_{1a}(x,y) = \begin{cases} 1 & \text{if } q_1(x) \geq 5 \wedge y = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{1b}(x,y) = \begin{cases} 1 & \text{if } q_1(x) \leq 5 \wedge y = 0 \\ 0 & \text{otherwise} \end{cases}$$

But one is really the "negation" of the other. So we can simply just have one such feature, let $f_1 := f_{1a}$. Next we can feature on how sleepy I am coming to the lecture.

$$f_2(x,y) = \begin{cases} 1 & \text{if } q_2(x) \geq 5 \wedge y = 1 \\ 0 & \text{otherwise} \end{cases}$$

A feature on whether the lecture was on zoom:

$$f_3(x,y) = \begin{cases} 1 & \text{if } q_3(x) = 0 \wedge y = 0 \\ 0 & \text{otherwise} \end{cases}$$

A feature on my background knowledge:

$$f_4(x,y) = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Let my feature vector be weight $w \in \mathbb{R}^k$. As of my experience, I almost get nothing out whenever its a zoom lecture. So I will give a high weight w_3. There should also be some weight to f_4 as I still get something out.
Here is another feature that I might consider. This feature is not *binary*. This i my cumulative knowledge increase as getting older:

$$f_5(x,y) = \begin{cases} q_5(x) & \text{if } y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Let me remark on how one can easily enlarge the number of features to hundreds.

- my scale can be finer, going from 1-100.

- I can vary my constraints on features.

$$f(x,y) = \begin{cases} 1 & \text{if } \bigwedge_{i=1}^{N} \{q_i(x) = a_i\} \wedge y = 1 \\ 0 & \text{otherwise} \end{cases}$$

  where $\{a_i\}_{i=1}^{N} \in \mathbb{R}$.

- Increase the value of N. In example, I have $N = 5$.

Training data

Clearly, these are based on my experience. Perhaps for better model, incorporating other student's data might help.

1.1 Word similarity

Question 8. here I used the `words-50.txt`. The most similar words to :
`seattle`

```
    dallas atlanta wichita tacoma lauderdale florida spokane chino
       dulles
```

`dog`

```
    badger cat hound puppy dachshund sighthound poodle rat keeshond
```

`jpg`

```
    svg szczepanek buteo pix gif image galleria regnum fiav
```

`the`

```
    its of which entire within from a part second
```

`google`

```
    word not in vocabulary
```

Below I copied for `words-20.txt` and `words-100.txt` for `the`.

```
    in within its between entire over part uninterrupted marked
```

```
    its in which entire a itself this second from
```

For larger values of d , the result seem closer to what we think of similarity.
Now let us make some additions:

```
python3 findsim.py words-50.txt king --minus man --plus woman
```

```
    queen throne carloman son melisende disgrace sibylla daughter
       betrothed
```

```
python3 findsim.py words-50.txt hitler --minus germany --plus italy
```

```
    cesare petacci innitzer banality honoria accomplices benito
       conspirators aetius
```

Some tests

Example 1: abstract realtion *worked ok.*

```
python3 findsim.py words-50.txt love --minus heart --plus brain
```

```
    emotion feelings ashamed senses thoughts unrequited imagining
       intellect hypnotist
```

Example 2: job-relations *not really worked.*

```
    python3 findsim.py words-200.txt teacher --minus school --plus
       doctor
```

```
    davros contemplating daleks yueh firefly mcgann pangloss marple
       scifi
```

Example 3: concrete relations which do not work

```
python3 findsim.py words-50.txt meow --minus cat --plus bark
```

On `words-50.txt`

```
    tablecloth pomegranates bark quenched dried nisibis sinai fayyum
       groves
```

and on `words-100.txt`

```
    willow trees pine alder sap munching poplar hardwoods quercus
```

and on `words-200.txt`

```
    excelsa pine trees leaf olive poplar leaves bushes ebony
```

Example 4: people-relation *Worked.*

```
python3 findsim.py words-100.txt teacher --minus student --plus boss
```

```
    henchmen stooge karras realises kingpin riddler thief dugan
       vasquez
```

Discussion:

- I was expecting things where relation is more "concrete" to work better. This wasn't the case for `meow-cat+bark`.
- In some cases, as example 2,3 Increasing from 50- 200 did not help.