Naïve Bayes classifier

Positive or negative movie review?

- This movie is disappointing.
- This movie does not disappoint.
- I would love to have that two hours of my life back.
- This is one of my favorite if not favorite films.
- I have seen so many bad low budget films lately, but I love this one.

Probabilistic model

Deterministic model

$$f(\text{I love, love this movie.}) = 0 \text{ or } 1$$

Probabilistic model

$$\begin{aligned} &f(\text{I love, love this movie.}) \\ &\quad = \mathbb{P}(y = 1 | \text{I love, love this movie.}) \end{aligned}$$

Naïve Bayes

- Simple (naïve) classifier based on Bayes rule:
  For a document $d = \{w_1, w_2, \ldots, w_n\}$ and a class $y = 0$ or 1

$$\mathbb{P}(y = 1 | w_1, w_2, \ldots, w_n) = \frac{\mathbb{P}(w_1, w_2, \ldots, w_n | y = 1)\mathbb{P}(y = 1)}{\mathbb{P}(w_1, w_2, \ldots, w_n)}$$

$$\mathbb{P}(y = 0 | w_1, w_2, \ldots, w_n) = \frac{\mathbb{P}(w_1, w_2, \ldots, w_n | y = 0)\mathbb{P}(y = 0)}{\mathbb{P}(w_1, w_2, \ldots, w_n)}$$

- Compute $\mathbb{P}(w_1, w_2, \ldots, w_n | y)$ and $\mathbb{P}(y)$ from the (labeled) data.

- What about $\mathbb{P}(w_1, w_2, \ldots, w_n)$?

Conditional independence

We assume that $\mathbb{P}(w_i|y)$ are independent given the class y.

$$\mathbb{P}(w_1, w_2, \ldots, w_n|y) = \mathbb{P}(w_1|y)\mathbb{P}(w_2|y)\ldots\mathbb{P}(w_n|y).$$

Graphical model

Maximum a posteriori (MAP)

$$\mathbb{P}(w_1, w_2, \dots, w_n|y)\mathbb{P}(y) = \mathbb{P}(y)\mathbb{P}(w_1|y)\mathbb{P}(w_2|y)\dots\mathbb{P}(w_n|y)$$

where

$$\mathbb{P}(w_i|y) = \frac{count(w_i, y)}{\sum_{w \in V} count(w, y)}$$

and

$$\mathbb{P}(y) = \frac{countdoc(Y = y)}{count(\text{Documents})}$$

Example

- This movie is disappointing.
- This movie does not disappoint.
- I would love to have that two hours of my life back.
- This is one of my favorite if not favorite films.
- I have seen so many bad low budget films lately, but I love this one.

$$\mathbb{P}(\text{Positive}) =$$

Example

- This movie is disappointing.
- This movie does not disappoint.
- I would love to have that two hours of my life back.
- This is one of my favorite if not favorite films.
- I have seen so many bad low budget films lately, but I love this one.

$$\mathbb{P}(\text{favorite}|\text{Positive}) =$$

Example

- This movie is disappointing.
- This movie does not disappoint.
- I would love to have that two hours of my life back.
- This is one of my favorite if not favorite films.
- I have seen so many bad low budget films lately, but I love this one.

$$\mathbb{P}(\text{disappoint}|\text{Positive}) =$$

Example

$$\mathbb{P}(y = 1 | \text{I love, love this movie.})$$

Example

$$\mathbb{P}(y = 0 | \text{I love, love this movie.})$$

Laplace smoothing

- What if there is a word **slept** in the test set, but not in the training set?

$$\mathbb{P}(\text{slept}|y) = \frac{count(\text{slept}, y)}{\sum_{w \in V} count(w, y)} = 0.$$

- There is no best y in this case.

$$\mathbb{P}(y|\text{slept}, \ldots) = \mathbb{P}(y)\mathbb{P}(\text{slept}|y) \ldots = 0$$

Laplace smoothing

Fix $\alpha > 0$.

$$\mathbb{P}(w_i|y) = \frac{count(w_i, y) + \alpha}{\sum_{w \in V}(count(w, y) + \alpha)}$$

$$= \frac{count(w_i, y) + \alpha}{\sum_{w \in V} count(w, y) + \alpha|V|}$$

In the previous example, if we choose $\alpha = 1$,

$$\mathbb{P}(\text{slept}|y) = \frac{\alpha}{\sum_{w \in V} count(w, y) + \alpha|V|}.$$

Learning Naïve Bayes

- From the training corpus, extract the **Vocabulary**.
- For each class y, calculate $\mathbb{P}(y)$
  - Count number of documents in class y.
  - $\mathbb{P}(y) = \frac{countdoc(Y=y)}{count(\text{Documents})}$
- For each word w_i and class y
  - Merge all documents in class y
  - $n_i \leftarrow$ # of occurrence of each word in class y
  - $\mathbb{P}(w_i|y) = \frac{n_i+\alpha}{\sum_i n_i+\alpha|Vocab|}$