

Machine Learning Review

Naïve Bayes

Bayes' rule applied to documents and classes

- For a document **d** and a class **c**

$$P(c | d) = \frac{P(d | c) P(c)}{P(d)}$$

$$\begin{aligned} C_{map} &= \operatorname{argmax} P(c | d) \\ &= \operatorname{argmax} \frac{P(d | c) P(c)}{P(d)} \\ &= \operatorname{argmax} P(d | c) P(c) \end{aligned}$$

Bayes classifier

$$\begin{aligned} C_{map} &= \operatorname{argmax} P(d | c) P(c) \\ &= \operatorname{argmax} P(x_1, x_2, \dots, x_{n.} | c) P(c) \end{aligned}$$

Documents represented
as features

- Bag of words assumption: assume position doesn't matter.
- Conditional independence: assume feature probabilities $P(x_i | c_j)$ are independent given class c .

$$P(x_1, x_2, \dots, x_{n.} | c) = P(x_1 | c) * P(x_2 | c) \dots P(x_n | c)$$

Bayes classifier

Positions= all words positions in the document

$$C_{NB} = \underset{j}{\operatorname{argmax}} P(c_j) \prod_{i \in \text{positions}} P(x_i | c_j)$$

because $\log(ab) = \log(a) + \log(b)$

$$C_{NB} = \underset{j}{\operatorname{argmax}} [\log P(c_j) + \sum_{i \in \text{positions}} \log P(x_i | c_j)]$$

Learning the Multinomial Naive Bayes Model

Maximum likelihood estimates

- Simply use the frequencies in the data

$$\hat{P}(c_j) = \frac{\text{doccount}(C=c_j)}{N_{doc}}$$

$$\hat{P}(w_i | c_j) = \frac{\text{count}(w_i, c_j)}{\sum_{w \in V} \text{count}(w, c_j)}$$

fraction of times word w_i appears among all words in documents of topic c_j

Smoothing Naive Bayes

What if your test data contains a word that is not in your training data?

$$\hat{P}(w_i | c_j) = \frac{\text{count}(w_i, c_j) + 1}{\sum_{w \in V} (\text{count}(w, c_j) + 1)}$$

$$= \frac{\overset{\text{某个字在文档中出现的次数}}{\text{count}(w_i, c_j) + 1}}{\underset{\text{一个文档中的所有字数}}{(\sum_{w \in V} \text{count}(w, c_j)) + |V|}}$$

Default probability
of every word is 1
over vocabulary