Machine Learning Review

Naïve Bayes

Bayes' rule applied to documents and classes

- For a document d and a class c

$$P(c \mid d) = \frac{P(d \mid c)\, P(c)}{P(d)}$$

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

Bayes classifier

$$C_{map} = argmax\ P\ (d \mid c)\ P\ (c)$$
$$= argmax\ P\ (\ x_1, x_2, \dots \dots x_{n.} \mid c\ )\ P\ (c)$$

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.

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

Bayes classifier

Positions= all words positions in the document

$$C_{NB} = argmax\, P\left(c_j\right) \prod_{i\, \in positions} P\left(x_i \mid c_j\right)$$

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

$$C_{NB} = argmax\, [\log P\left(c_j\right) + \sum_{i\, \in positions} \log P\left(x_i \mid c_j\right)]$$

Learning the Multinomial Naive Bayes Model

Maximum likelihood estimates

- Simply use the frequencies in the data

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

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

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?

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

$$= \frac{count\,(w_i, c_j)+1}{\left(\sum_{w \in V} count(w, c_j)\right)+|V|}$$

Default probability of every word is 1 over vocabulary