## 1.Naive Bayes and Logistic Regression

1.generative classifiers: E.g 2 models , 1 picture, which one fits better

discriminative(判别) classifier: distinguish dogs from cats

2.Naive Bayes: generative linear classifier, use add-1 smooth

ignore the unknown words, use all words and dont use stopword lists

is the count of Cj,

is the count of all words of c, |V| is the count of vocabulary

=

3.Logistic Regression: discriminative linear classifier



4.In text classification, the potential drawbacks of using a very large number of features: overfitting, underfitting, slower training speed

5.factors affect the performance of a text classification model: Quality of training data , Choice of classification algorithm , Size of the vocabulary

## 2.Neural Network

1.Activation Functions: Sigmoid: 图示

描述已自动生成 derivative=y(1-y)

tanh: 手表上有字

描述已自动生成 ReLu: 

2.steps in training a neural network for NLP tasks: Parameter Initialization , Forward Pass Backward Pass

## 3.POS tagging(词性标注) and NER(命名实体识别)

1.closed class word: usually function words, relatively fixed

open class word: content words

2. tag accuracy in English is bout 97%

3. Information Sources for POS tagging: Prior probabilities of word/tag, Identity of neighboring words, Morphology (wordshape)

4.the main task of NER: Identifying and classifying named entities in text.

5.BIO tagging: B: token that begins a span I: tokens inside a span O: tokens outside of any span.

6.algorithms for NER: Hidden Markov Models, Conditional Random Fields (CRF)/ Maximum Entropy Markov Models (MEMM), Neural sequence models (RNNs or Transformers), Large Language Models (like BERT), finetuned

7. Both POS tagging and NER are based on sequence tagging (word-level classification).

## 4. Sentiment Analysis

1. Binary multinomial Naïve Bayes clips word counts at 1, which is simple yet effective.

2. the simple baseline method dealing with negation is adding NOT\_ to every word between negation and following punctuation

3.lexicon:

1). MPQA Subjectivity Cues Lexicon: 6885 words from 8221 lemmas, annotated for intensity (strong/weak) and(p/n)

2). The General Inquirer: Positive and Negative, Strong vs Weak, Active vs Passive, Overstated vs Understated, Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation

4. Two families of theories of emotion: Atomic basic emotions, Dimensions of emotion: Valence (positive negative), Arousal (strong, weak), Dominance (the degree of control)

5. Affective Analysis(a broader field)

NRC Word-Emotion Association Lexicon labeled by multiple emotions with different intensities.

6. The 2-by-2 confusion matrix: don’t use accuracy. A dumb classifier can achieve high accuracy by simply labeling everything as negative, which is not useful for finding positive instances.

Precision: tp/(tp+fp) 系统检测的事实上positive的比例

Recall: tp(tp+fn) 输入中被正确区分的比例

F1=2PR(P+R)

7. Cross-validation : Train on training set, tune on devset(Development Test Set), report on testset, multiple splits, Pool results over splits, Compute pooled dev performance(汇总devset上的结果进行评估)

8. Macroaveraging：compute performance, then average.

Microaveraging: collect all decisions into one confusion matrix, then compute