

- Classification is the task of choosing the correct class label for a given input.
- A classifier is called supervised if it is built based on training corpora containing the correct label for each input.
- the author start to create simple classifier to classify between male and female
- Selecting relevant features and deciding how to encode them for a learning method can have an enormous impact on the learning method's ability to extract a good model.
- The training set is used to train the model, and the dev-test set is used to perform error analysis. The test set serves in our final evaluation of the system
- One shortcoming of this approach is that we commit to every decision that we make. For example, if we decide to label a word as a noun, but later find evidence that it should have been a verb, there's no way to go back and fix our mistake. One solution to this problem is to adopt a transformational strategy instead
- Transformational joint classifiers work by creating an initial assignment of labels for the inputs, and then iteratively refining that assignment in an attempt to repair inconsistencies between related inputs.
- Another solution is to assign scores to all of the possible sequences of part-of-speech tags, and to choose the sequence whose overall score is highest. This is the approach taken by Hidden Markov Models
- . Hidden Markov Models are similar to consecutive classifiers in that they look at both the inputs and the history of predicted tags
- Rte is task used to find the meaning of paragraph in just short sentence
- recap of test set , accuracy , confusion matrix , precision , recall and f1 score
- In order to evaluate our models, we must reserve a portion of the annotated data for the test set. As we already mentioned, if the test set is too small, our evaluation may not be accurate. However, making the test set larger usually means making the training set smaller, which can have a significant impact on performance if a limited amount of annotated data is available.
- A second, and equally important, advantage of using cross-validation is that it allows us to examine how widely the performance varies across different training sets. If we get very similar scores for all N training sets, then we can be fairly confident that the score is accurate. On the other hand, if scores vary widely across the N training sets, then we should probably be skeptical about the accuracy of the evaluation score.
- recap on Naive bayes , decision tree and entropy classifier
- generative models are strictly more powerful than conditional models, since we can calculate the conditional probability  $P(\text{label}|\text{input})$  from the joint probability  $P(\text{input}, \text{label})$ , but not vice versa
- In naive Bayes classifiers, each feature independently contributes to the decision of which label should be used. This allows feature values to interact, but can be problematic when two or more features are highly correlated with one another.
- Maximum Entropy classifiers use a basic model that is similar to the model used by naive Bayes; however, they employ iterative optimization to find the set of feature weights that maximizes the probability of the training set.
-