1. Features will change as training sets change. What exactly are our features?
2. How to deal with coauthors?
3. For algorithm A, it is said that A is clustering algorithm. So why do we need ^ here?
4. In error-driven online training, whether the "first error" means the first T' in the neighbor of T(which we define T' as N\*(T)) s.t. S\*(T') > S\*(\hat T). When error occurs, we will update the para(t) --> para(t+1), 然后接下来要去执行哪步？
5. Meaning of X
6. F in perceptron meaning?
7. Score function: p1 & p?
8. Prediction algorithm: can use any clustering method? K-means? (how to know the value of k beforehand?) Agglomerative clustering? (can we use any packages?)