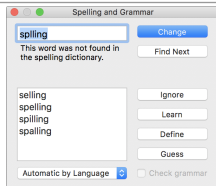


What are...nearest neighbors?

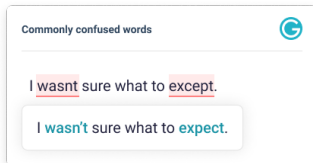
Or: Close friends

Data clusters of words

this:

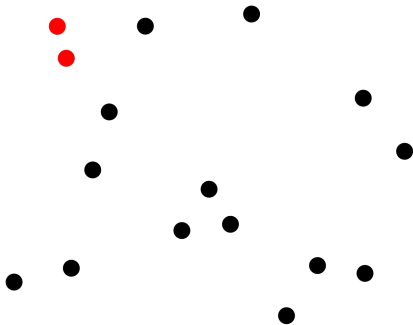


not this:



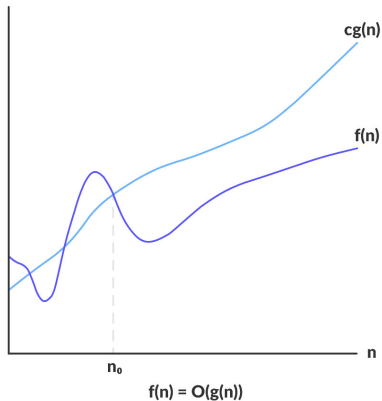
- ▶ How does a **spell checker** work?
- ▶ **Simplified version**
 - ▷ We have a data base of words encoded as points in some \mathbb{R}^n
 - ▷ The word we want to check is another point
 - ▷ Then we look for the **closest neighbor** point

The geometric reformulation



-
- ▶ **Problem** For n points in \mathbb{R}^d , find the pair that is closest
 - ▶ The meaning of “closest” varies on the problem (e.g. this problem works in any metric space)
 - ▶ **Question** How to find the closest pair efficiently?

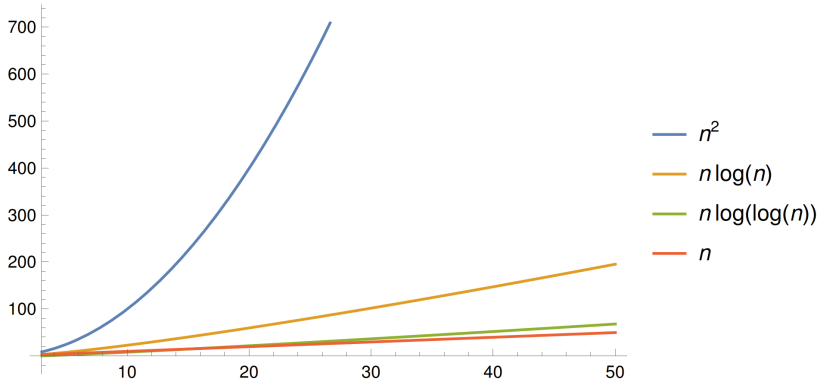
A naive approach



- ▶ Recall $\sum_{k=1}^n k = n(n-1)/2 \approx n^2$
- ▶ This naive algorithm – measure all distances – runs in $O(n^2)$
- ▶ Big O notation “Is bounded from above” see illustration

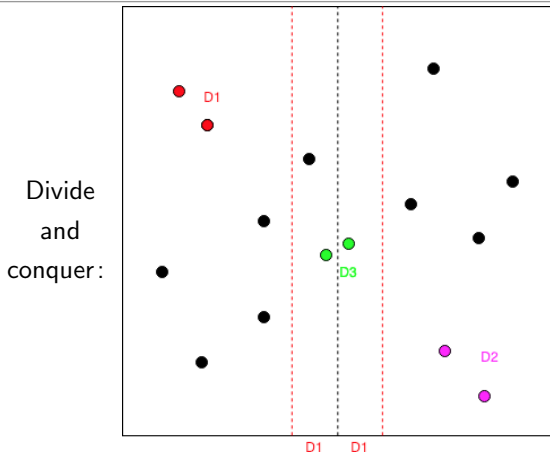
Enter, the theorem

Closest pair of points is in $O(n \log \log n)$



- ▶ $n \log \log n$ is essentially linear
- ▶ Randomized algorithms can even solve this in linear time
- ▶ In the above we have the dimension d fixed

An $n \log n$ algorithm



- ▶ **Divide** Cut the problem in half and solve each half
- ▶ **Safe time** by only focusing on points within a region around the cut
- ▶ **Conquer** Do this recursively

Thank you for your attention!

I hope that was of some help.