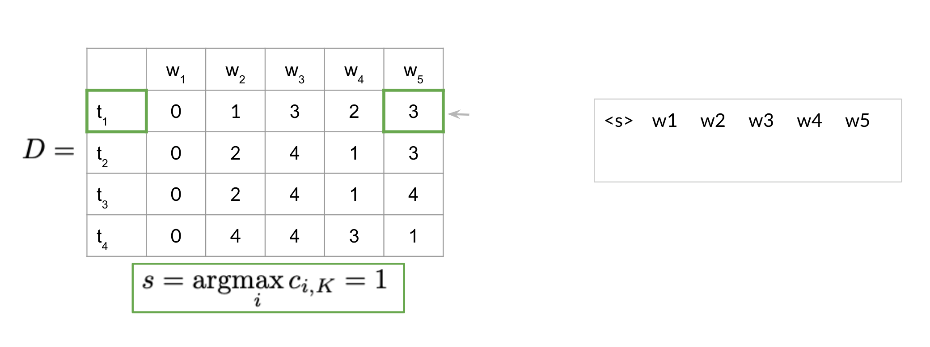
## Week 1 - Auto-Correct and Minimum Edit Distance

1. The minimum edit distance between the words ‘deep’ and ‘creepy’ is:
   1. 4 (You need to replace *d* for *c*, which counts for 2, insert *r* and insert *y*.)
2. What is NOT a valid example of an edit string operation?
   1. SWITCH a letter ‘Lusca’ --> ‘Lucas’ (Switching a letter is a valid operation ONLY when switching adjacent letters. In this case, there were two switches: *s* and *c*, after *s* and *a*.)
3. Autocorrect is only applicable when dealing with misspelled words.
   1. False. (Autocorrect can be used for words that do not make any sense for a particular sentence. For example, ‘Happy birthday deer friends’ is a correctly-spelled sentence, but the word ‘deer’ makes no sense – it should be ‘dear’.)
4. Given the corpus: “I am happy because I am doing quizzes.” and the following sentence: “I sm very good at solving quizzes.”, what is true?
   1. There is a unique correction for the misspelled word “sm”. (The correction would be the word “am”.)
5. What is true about the probabilistic model:
   1. Replacing a character costs more than deleting a character. (Replacing a word costs 2 whereas deleting it costs 1.)
   2. If is the number of times a word appears in a corpus and is the corpus size, then the probability of the word in the corpus is .
   3. The sentence “Happy birthday deer friends” would not have any word corrected. (Since the probabilistic model just looks at misspelled words, the above sentence would not be corrected.)
6. If we build a distance matrix for the following case: Source: Pie --> Target: Bye, what is the value of ?
   1. 5
7. If is the distance matrix for two words of the same size and is the matrix size for the minimum edit distance algorithm, then:
   1. (The first line will always have increasing values as we move to the right, because it is the cost from editing the null string.)
   2. The algorithm avoids use of brute force by implementing a dynamic programming approach. (Using previously computed cells to compute the next one is a dynamic programming method.)
8. What is NOT true about minimum edit distance?
   1. It is used to check if a word is misspelled.
9. The minimum edit distance calculation is more computationally expensive for a large corpus.
   1. False. (The minimum edit distance depends only on the editing cost and the two words that are being considered, not on any corpus or vocabulary.)
10. For the corpus “Autocorrect is a powerful tool and it is used on our computers”, the value for is:
    1. .

## Week 2 – Part of Speech Tagging

1. The transition matrix A allows you to:
   1. Compute the probability of going from a part of speech tag to another part of speech tag.
2. The emission matrix B allows you to:
   1. Compute the probability of going from a part of speech tag to a word.
3. The column sum of the emission matrix has to be equal to 1.
   1. False.
4. The row sum of the transition matrix has to be equal to 1.
   1. True.
5. Why is smoothing usually applied?
   1. Applying smoothing, for the minority of cases, allows us to increase the probabilities in the transition and emission matrices and this allows us to have non zero probabilities.
   2. Applying smoothing, for the majority of cases, allows us to decrease the probabilities in the transition and emission matrices and this allows us to have non zero probabilities.
6. Given the following D matrix, what would be the sequence of tags for the words on the right?



1. We had been multiplying the raw probabilities but in reality we take the log of those probabilities. Why?
   1. Because probabilities are bounded between 0 and 1 and as a result, the numbers could be too small and go toward 0.
2. What are useful applications for part of speech tagging?
   1. Speech recognition.
   2. Coreference resolution.
   3. Named entity recognition.

## Week 3 - Autocomplete

1. Corpus: “In every place of great resort the monster was the fashion. They sang of it in the cafes, ridiculed it in the papers, and represented it on the stage. ” (Jules Verne, Twenty Thousand Leagues under the Sea)

In the context of our corpus, what is the probability of word “papers” following the phrase “it in the”.

1. Given the following conditional probabilities, approximate the probability of the following sentence with bigrams: “Mary likes cats”:

; ; ; ;  
; ;

1. Given the following conditional probabilities, approximate the probability of the following sentence with bigrams:

; ; ; ;  
; ;

1. Given the logarithm of the following conditional probabilities, approximate the log probability of the following sentence with bigrams:

; ; ;

1. Given the same conditional probabilities, what is the model’s perplexity assuming the test set is
2. Given the following training corpus and a minimum word frequency of 2, what would the vocabulary for the corpus pre-processed with look like?

1. Given the following corpus, what is the estimates probability of the word “can” following the word “I” using the bigram model and adding k-smoothing where .

1. What are applications of n-gram language models?
   1. Speech recognition
   2. Auto-complete
   3. Auto-correct
   4. Augmentative communication
2. The higher the perplexity score, the more our corpus will make sense.
   1. False.
3. The perplexity score increases as we increase the number of tokens.
   1. False.