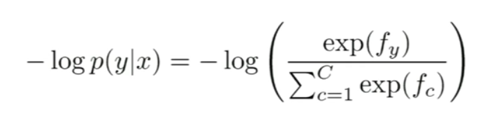
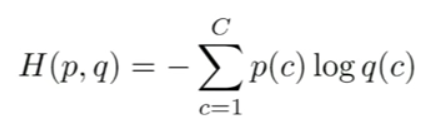
Lecture 4 | Word Window Classification and Neural Networks

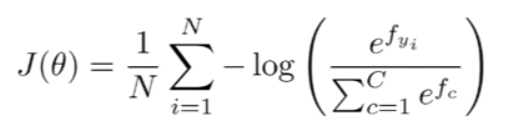
* Training with softmax and cross-entropy error
  + For each training example {x,y}, our objective is to maximise the probability of the correct class y
  + Therefore we want to minimise the negative log probability of that class:



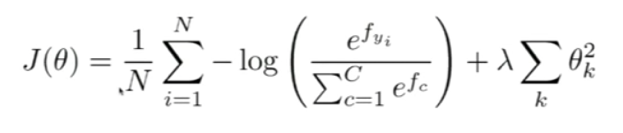
* + Why ‘Cross entropy’ error?
    - Assuming a ground truth probability distribution, p, that is 1 for the right class and 0 everywhere else (one-hot vector) and our computed probability is q, therefore the cross entropy is:



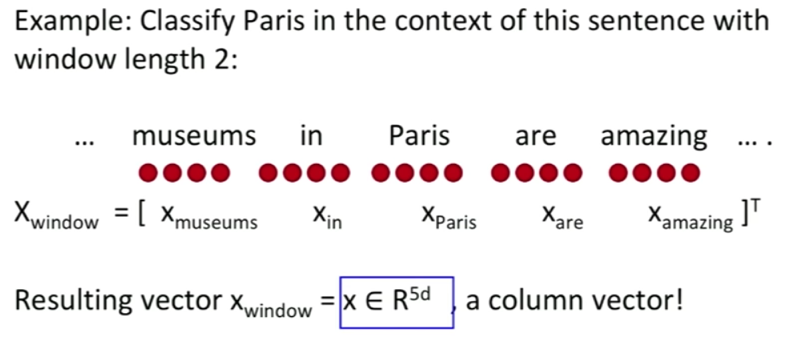
* + - * Because of one-hot p, the only term left is the negative log probability of the true class
  + Therefore, cross entropy loss function over full dataset is:



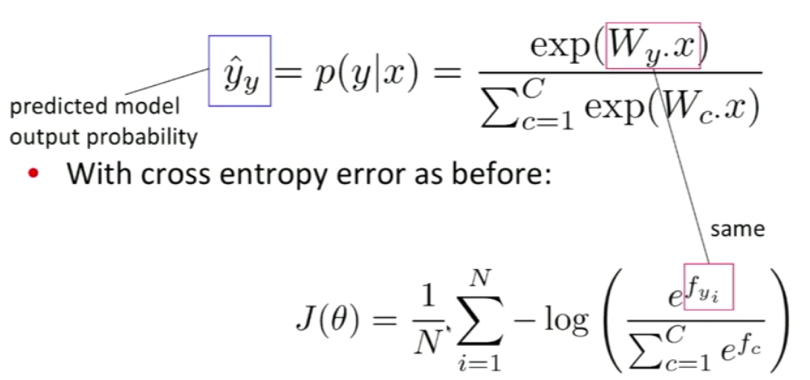
* + - With regularisation, you have:



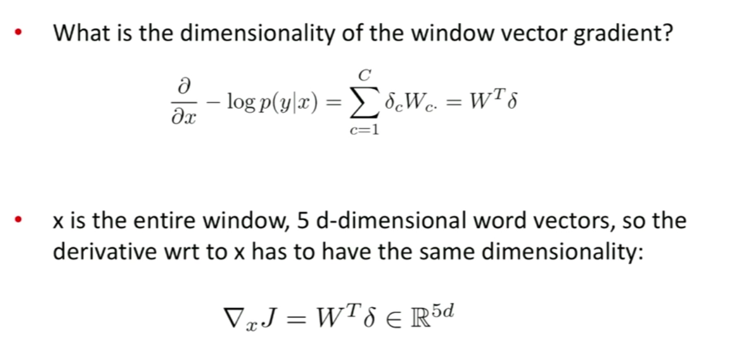
* + - * Where regularisation will prevent overfitting when we have lots of features
  + Losing generalisation by retraining word vectors
    - If you only have a small training data set, don’t train the word vectors
* Window classification
  + The idea is to classify a word in its context window of neighbouring words
  + For example, for every single word in our corpus, we want to categorise it into one of the 4 NER classes:
    - Person, location, organisation, none
  + Train a softmax classifier by assigning a label to a center word and concatenating all word vectors surrounding it. For example:

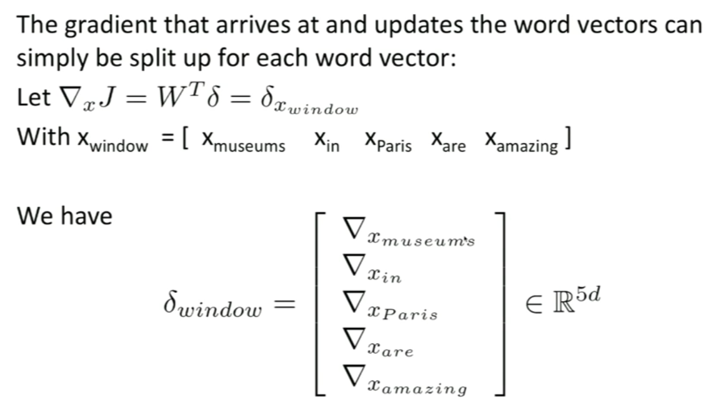


* + With x = x\_window, we can use the same softmax classifier and cross entropy formula as above



* + How do we update the word vectors?
    - Take derivatives (back prop) as before
    - Carefully define variables and keep track of their dimensionality
      * y\_hat : softmax probability output vector
      * t : target probability distribution (all 0’s except at ground trut index of class y, where its 1)
    - Use chain rule!
    - For softmax derivative, first take the derivative w.r.t. f\_c when c = y (correct class), then take derivative w.r.t. f\_c when c != y (all incorrect classes)
    - Use vector operations rather than if statements





* + Matrix implementations

