CRF: Conditional Random Fields

**Stochastic Gradient Descent:** This is an improvement over gradient descent method. Main advantage is uses lesser data to calculate gradient and computationally effective.   
Basically during calculating gradient descent rather than using the whole data set m (total size). This uses data set of size b (b<<m). If b==m, then it is conventional gradient descent. If b==1, then it is online learning.   
  
Convergence of SGD can be improved by *gain vector adaptation. Trick with this method is to update the gain vector (alpha) simultaneously, while updating theta.*