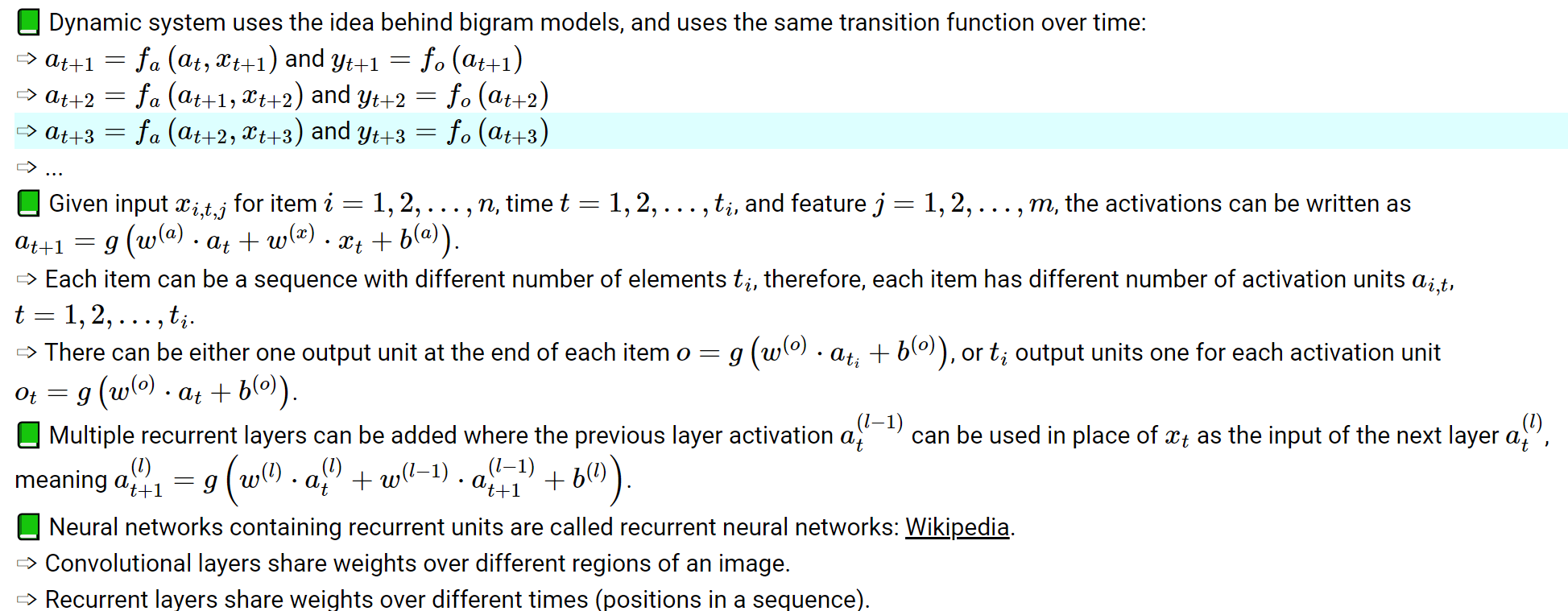
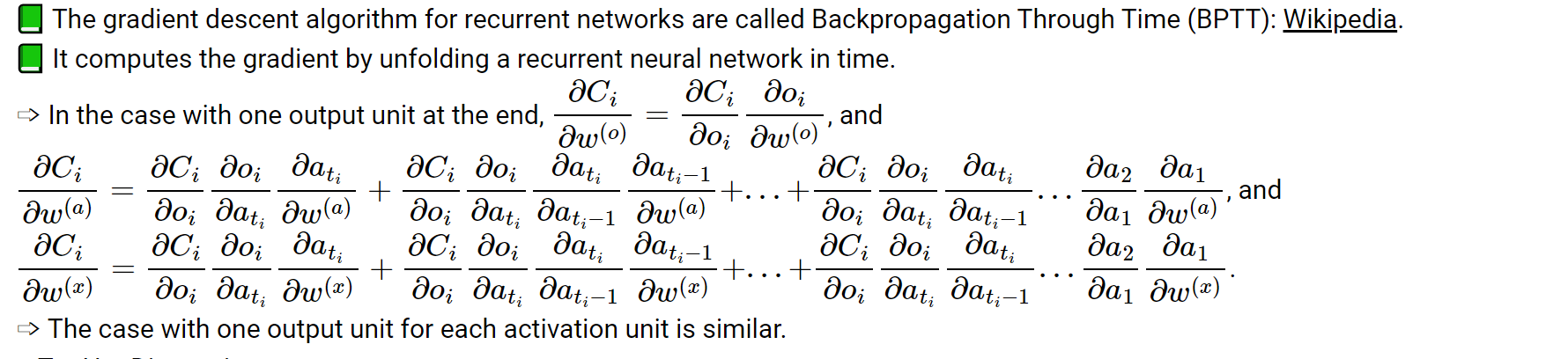
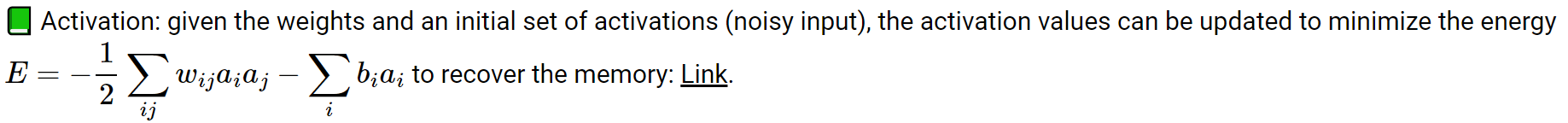
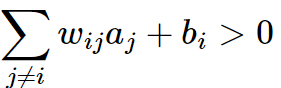
* Recurrent networks
  + 
* Bp through time
  + 
  + Repeat for each output if the output is a sequence
* Vanishing and exploding gradient
  + Vanishing gradient: if the weights are small, the gradient through many layers will shrink exponentially to 0
  + Exploding gradient: if the weights are large, the gradient through many layers will grow exponentially to ±∞.
  + In recurrent neural networks, if the sequences are long, the gradients can easily vanish or explode.
  + In deep fully connected or convolutional networks, vanishing or exploding gradient is also a problem
* Gated recurrent unit and long short term memory
* Hopfield networks
  + Used to store memories in local mins of the network
  + 
  + Energy function is the landscape that stores memory
  + Activation - ai is randomly chosen and updated to 1 if  and -1 otherwise
  + Training
    - Biases are not updated
* Generative adversarial network
  + Generator - input is noise, output is fake image
  + Discriminator -= input the fake or real image, output is binary class whether the images real
  + Same but opposite loss function - generator try to maximize the loss of the discriminator, the discriminator try to minimize the loss of classifying fake/real
* Diffusion
  + Train image to noise
  + Generate noise to image
* Transformers