



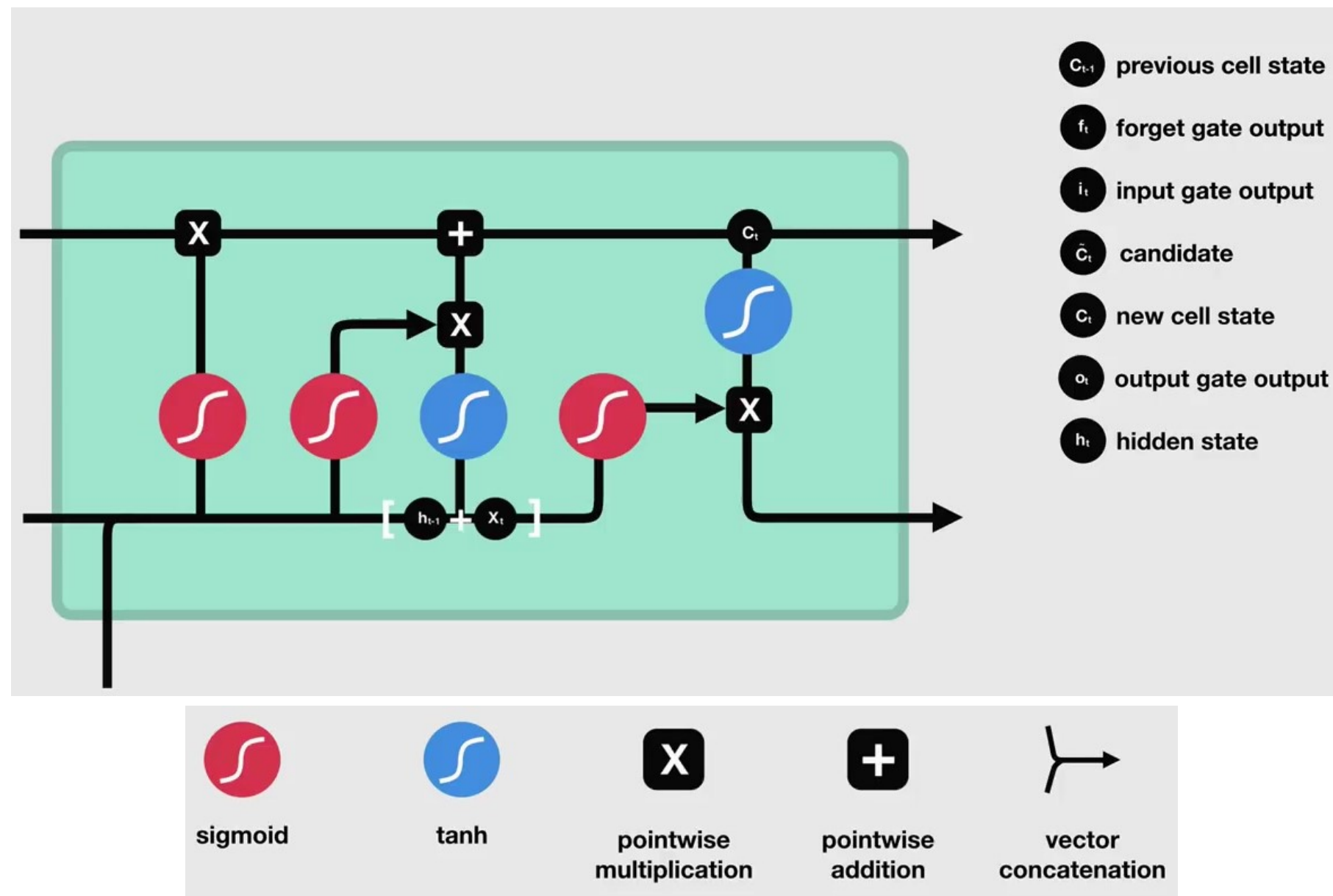
المحتويات

				التطبيقات	العقبات و التحديات	تاريخ NLP	ما هو NLP	المحتويات	(1) مقدمة
					البحث في النصوص	ملفات pdf	الملفات النصية	المكتبات	(2) أساسيات NLP
T.Visualization	Syntactic Struc.	Matchers	Stopwords	NER	Stem & Lemm	POS	Sent. Segm.	Tokenization	(3) أدوات NLP
	Dist. Similarity	Text Similarity	TF-IDF	BOW	Word2Vec	T. Vectors	Word embed	Word Meaning	(4) المعالجة البسيطة
T. Generation	L. Modeling	NGrams	Lexicons	GloVe	NMF	LDA	T. Clustering	T. Classification	(5) المعالجة المتقدمة
	Summarization & Snippets		Ans. Questions		Auto Correct	Vader	Naïve Bayes	Sent. Analysis	
Search Engine	Relative Extraction		Information Retrieval		Information Extraction		Data Scraping	Tweet Collecting	(6) تجميع البيانات
					Rec NN\TNN	GRU	LSTM	RNN	(7) RNN
Chat Bot	Gensim	FastText	Bert	Transformer	Attention Model	T. Forcing	CNN	Word Cloud	(8) تكتيكات حديثة

القسم السابع : الشبكات العصبية المتكررة

LSTM : الجزء الثاني

وهي اختصار Long Short-Term Memory

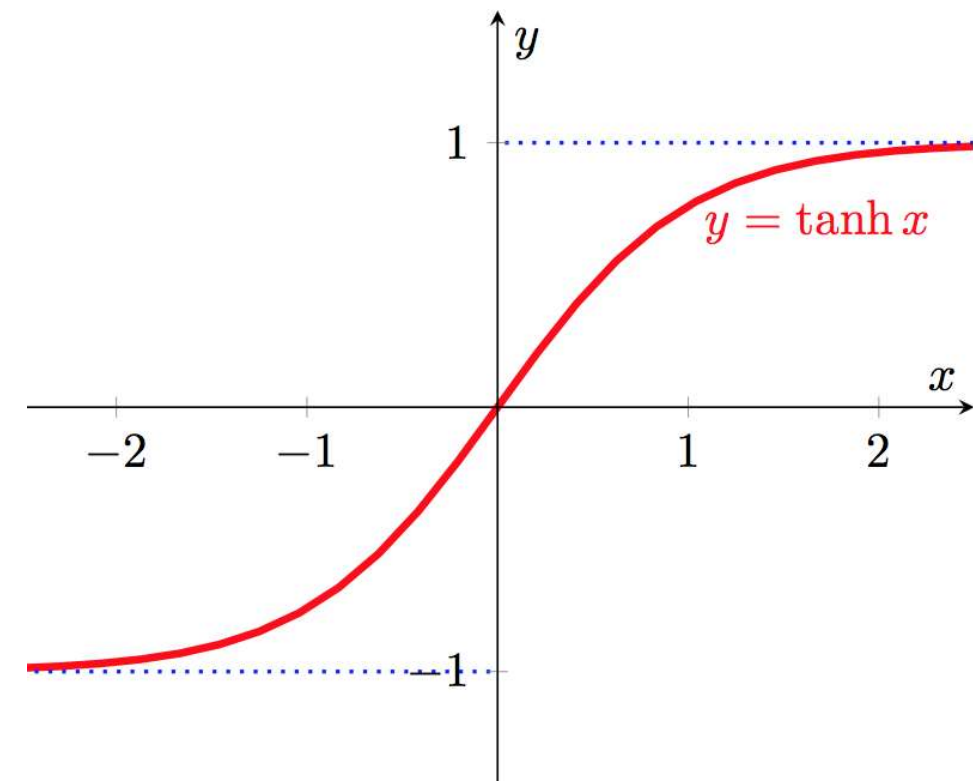
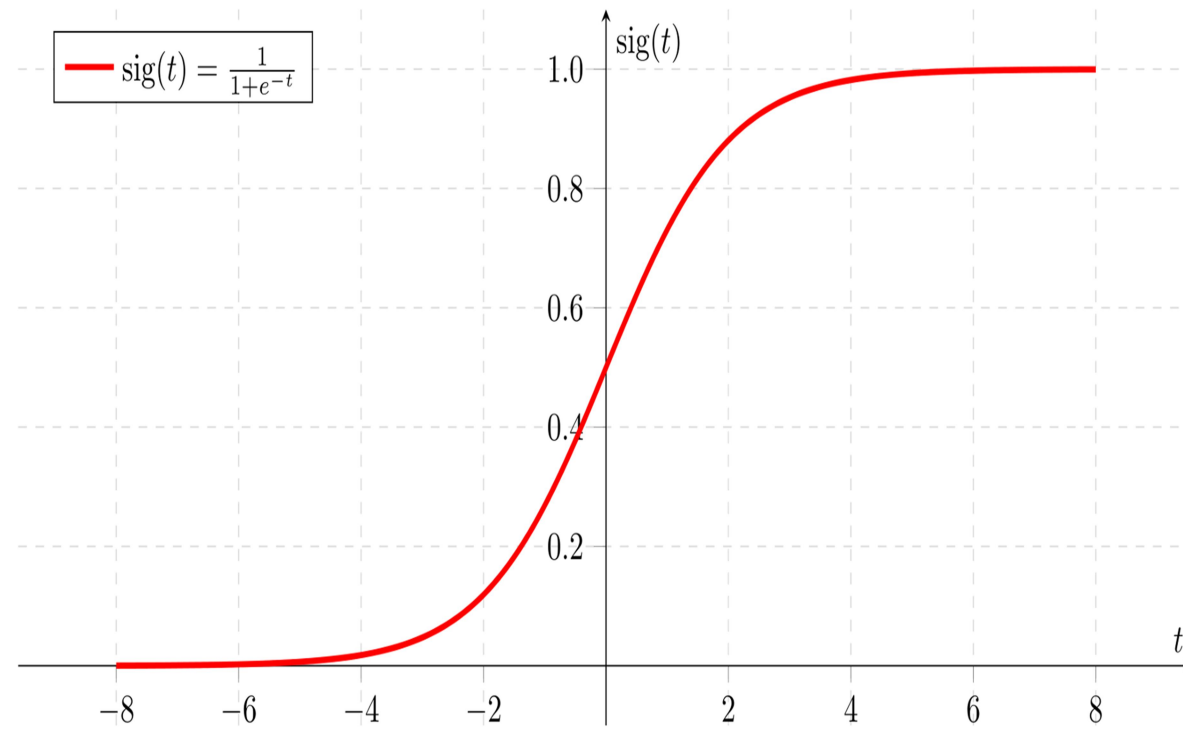


وهي تستخدم بشكل اساسي لمساعدة الموديل علي تذكر الكلمات الهامة في النصوص , والتركيز عليها بعيدا عن باقي الكلمات ذات الأهمية الأقل :

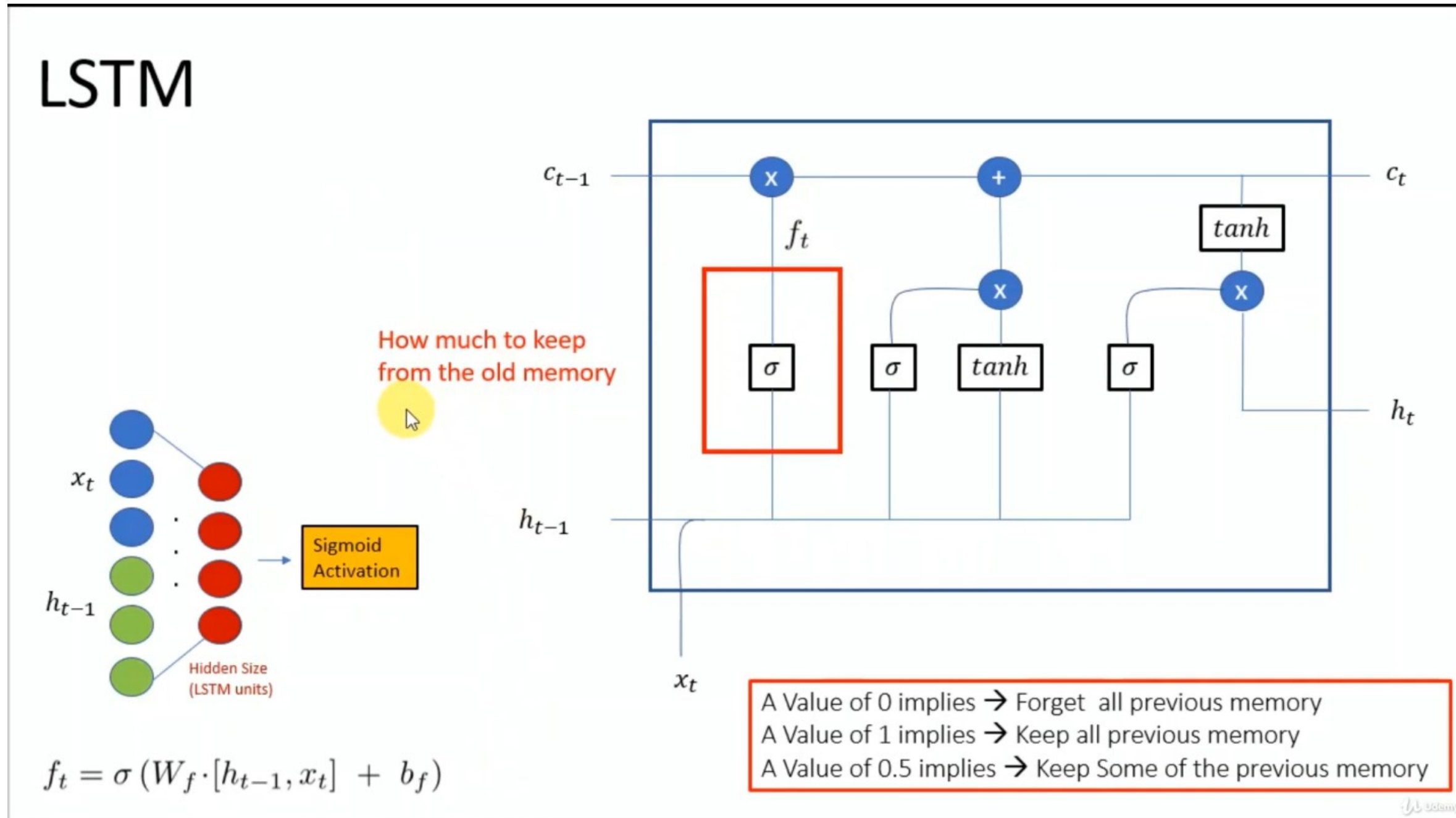
Amazing! This box of cereal gave me a perfectly balanced breakfast, as all things should be. I only ate half of it but will definitely be buying again!

Amazing! This box of cereal gave me a perfectly balanced breakfast, as all things should be. I only ate half of it but will definitely be buying again!

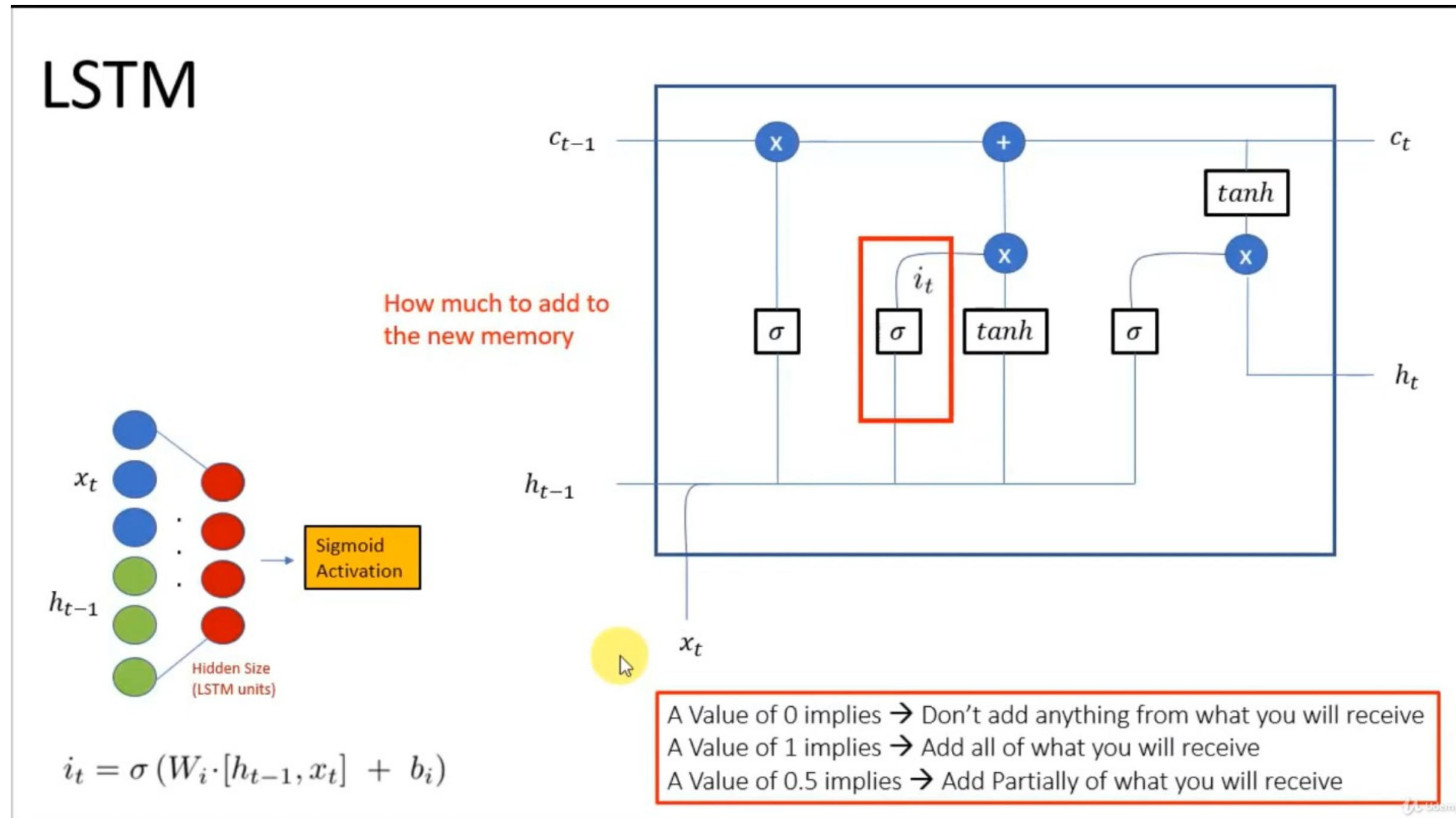
الفارق الأساسي بين دالة tanh , sigmoid هو المدي



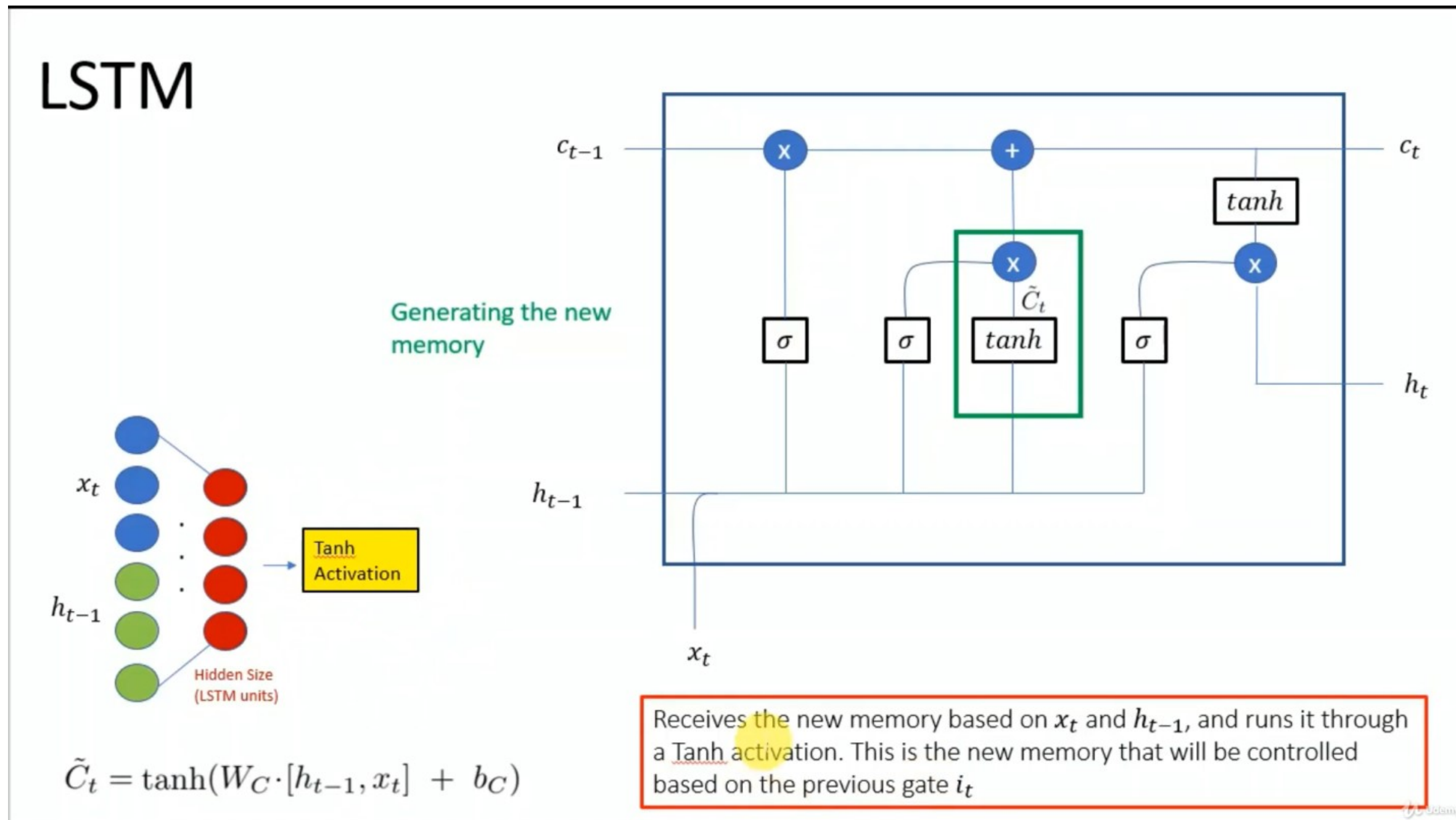
في الجزء الأول نري علامة السيجمما σ (الدلالة علي سيجمويد) وهي تمثل البوابات التي تتحكم في تدفق البيانات , وهي تتحكم في كمية البيانات الداخلة من الذاكرة القديمة , اي التي تأتي من ال Output السابق , واذا كانت بقيمة 0 او 1 او 0.5 فهي تتحكم في هذا المقدار من البيانات



أما الخلية التالية فهي تتحكم في مقدار البيانات التي سيتم اضافتها للخلية التالية , ولها قيم بقيمة 0 او 1 او 0.5 تدل علي مقدار الاضافة



و هنا نصل للخلية الثالثة : و هي التي تتناول البيانات الداخلة لها و تقوم بصناعة الذاكرة الجديدة

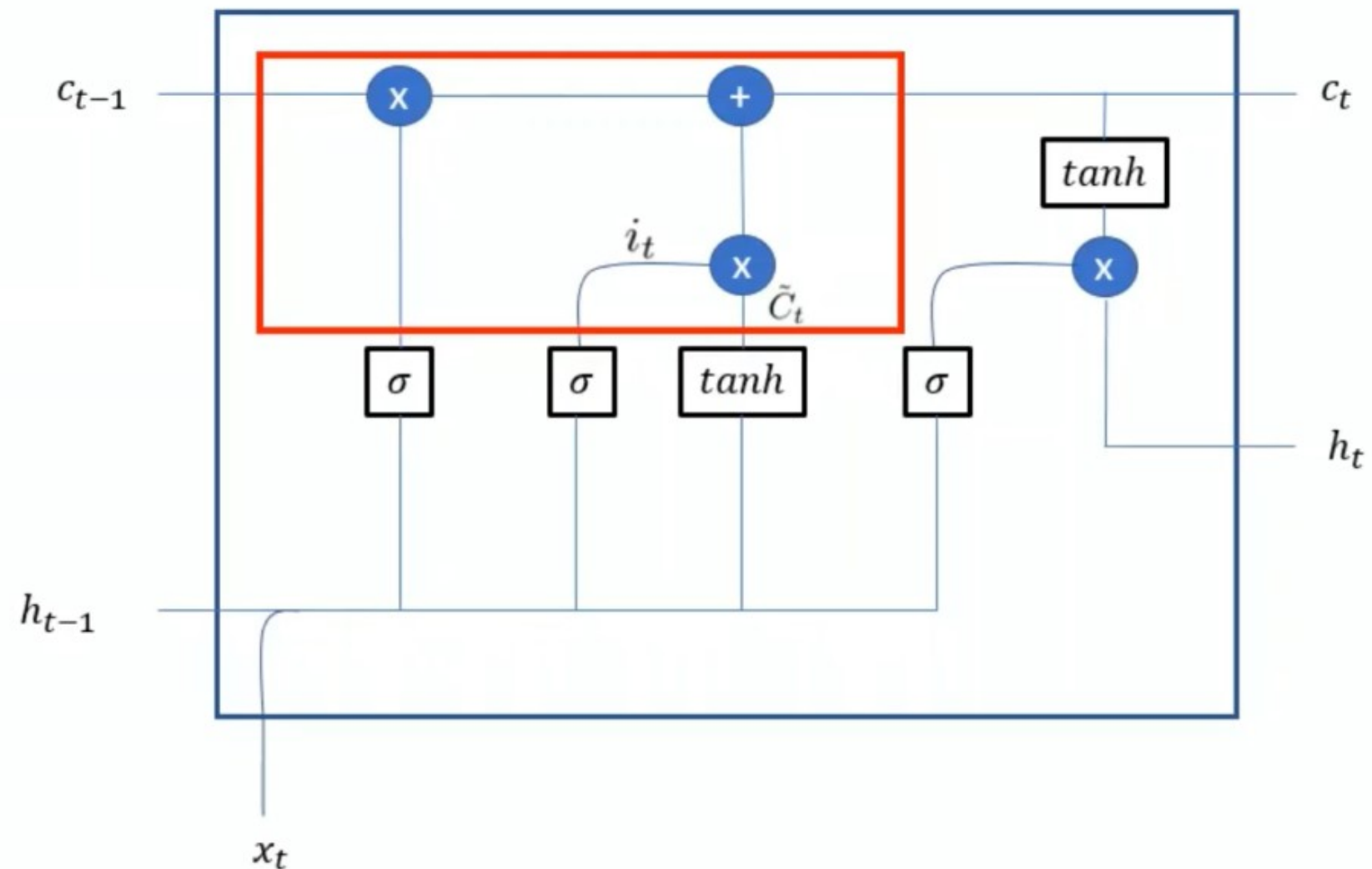


ثم نقوم بتصميم الذاكرة الجديدة للخلية بهذه المعادلة

LSTM

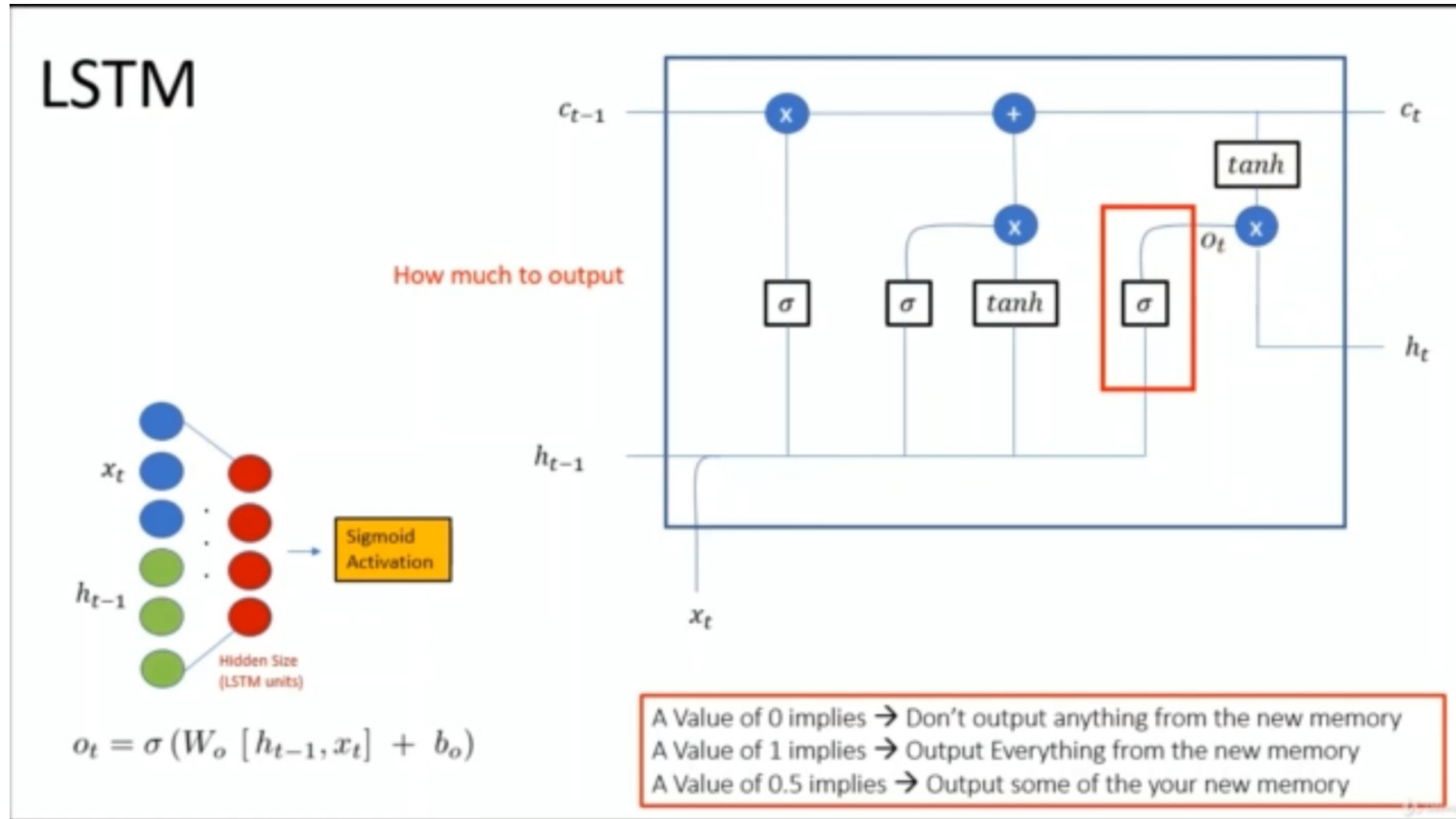
Forget what we don't need from the old memory, and add what we need from the new memory.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

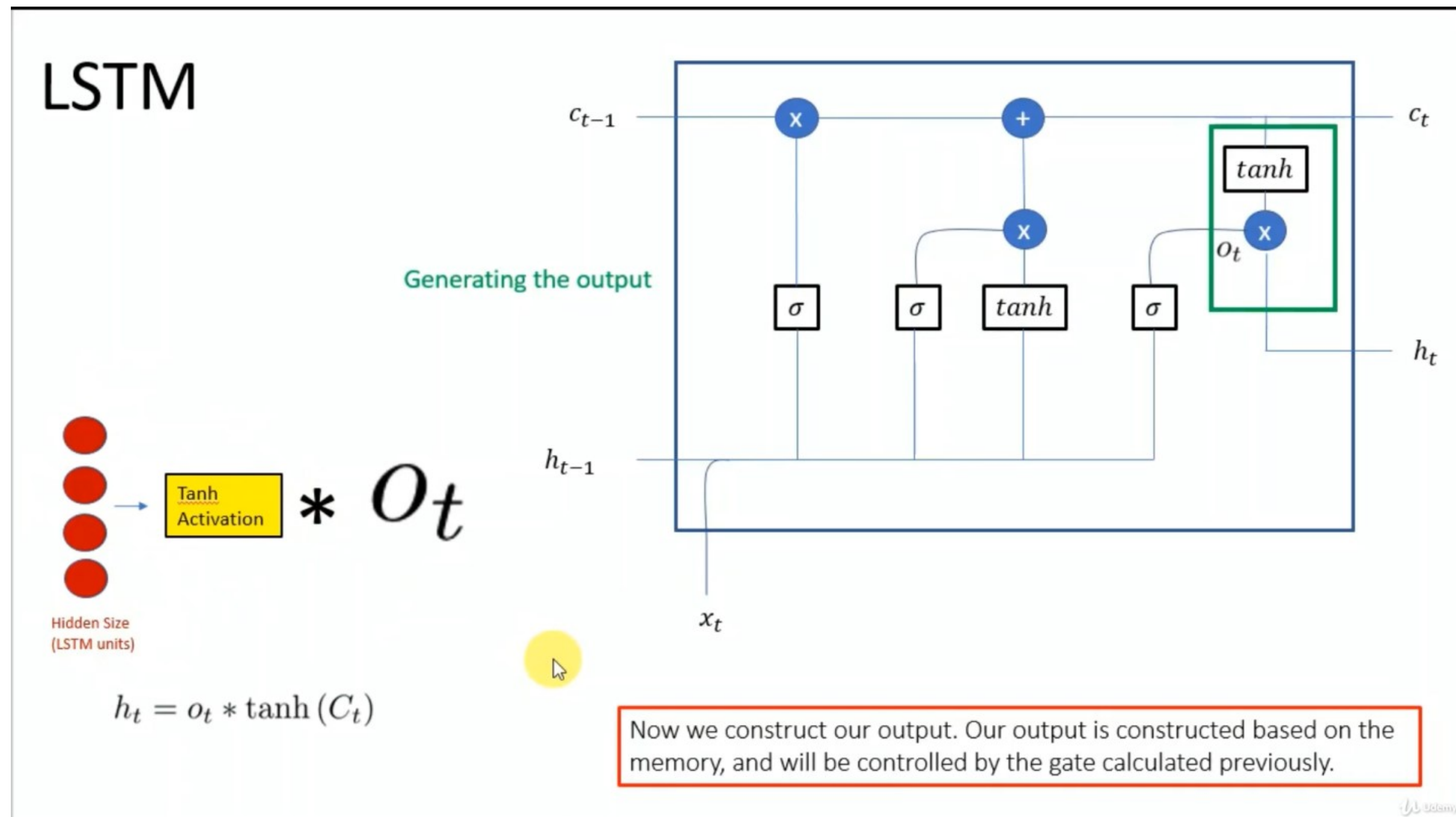


Now we construct our new memory

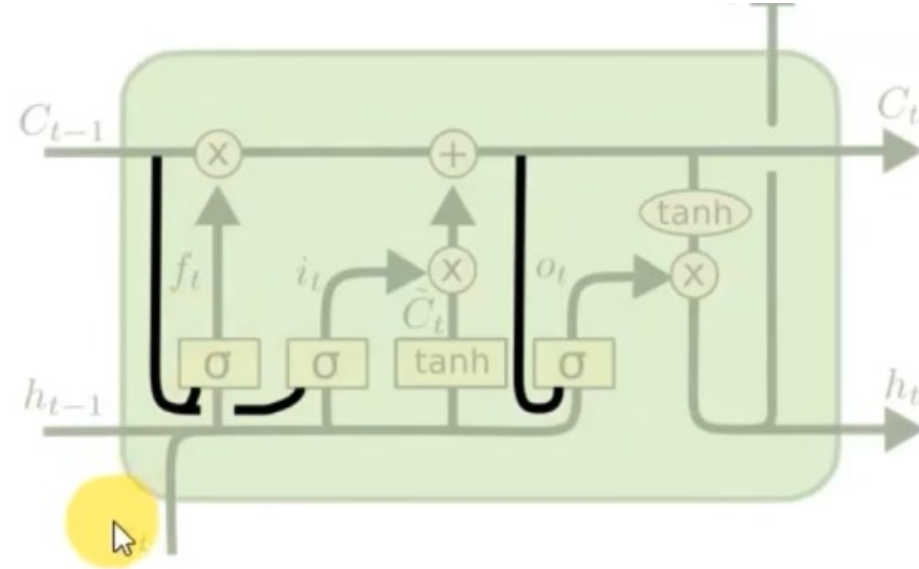
الخطوة التالية التحكم في كمية ال output من الخلية



ثم معالجة و انتاج ال output نفسه بمعادلة ال \tanh



و هناك بدائل للتصميم السابق , حيث يختلف الهيكل مما يسمح للشبكة ان تقوم بأداء مختلف مع الداتا المختلفة بهذا الشكل

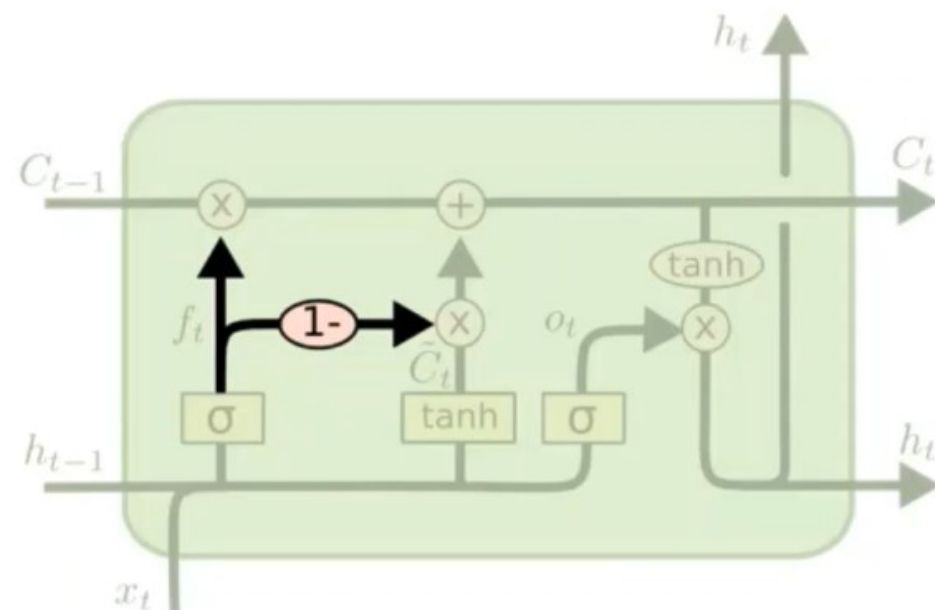


$$f_t = \sigma (W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma (W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma (W_o \cdot [C_t, h_{t-1}, x_t] + b_o)$$

و هذا الشكل



$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t$$

* _ * _ * _ * _ * _ * _ * _ * _ * _ * _ * _ * _ * _ *