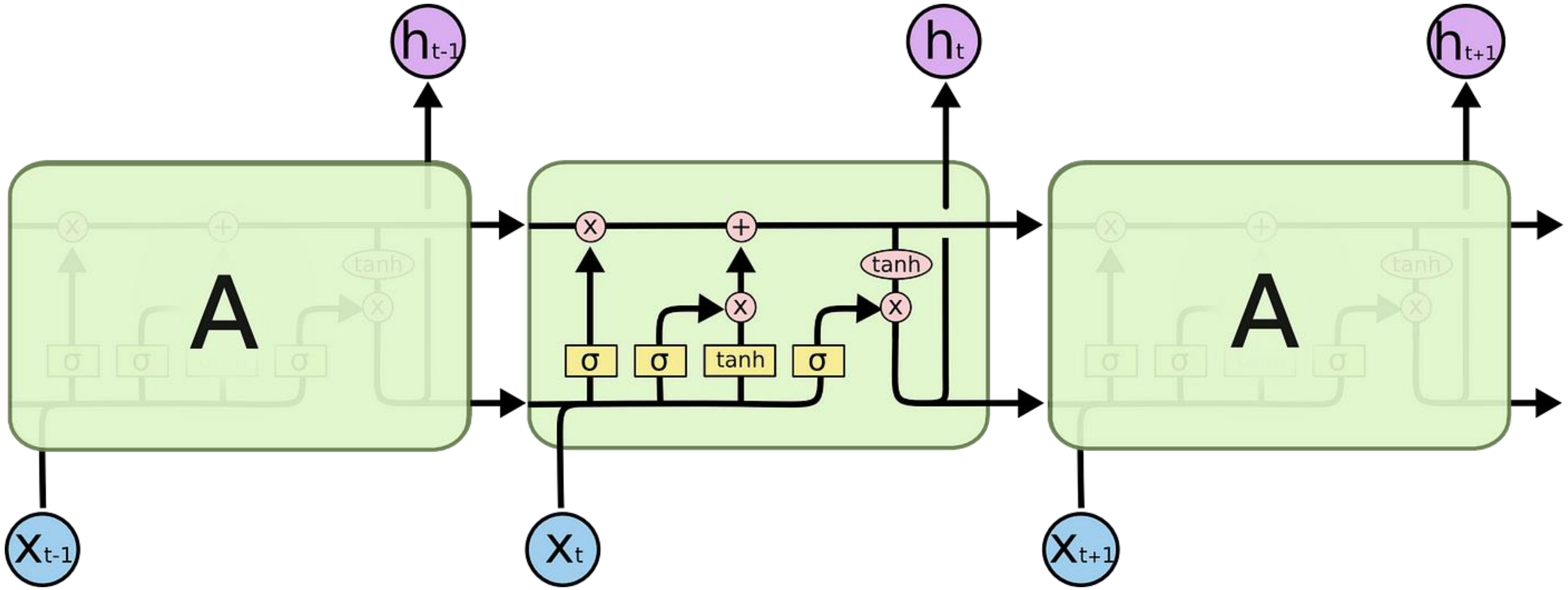
Spam or Not Spam Data Generation using LSTM VAE

Tejas Asija
Deevyansh Khadria



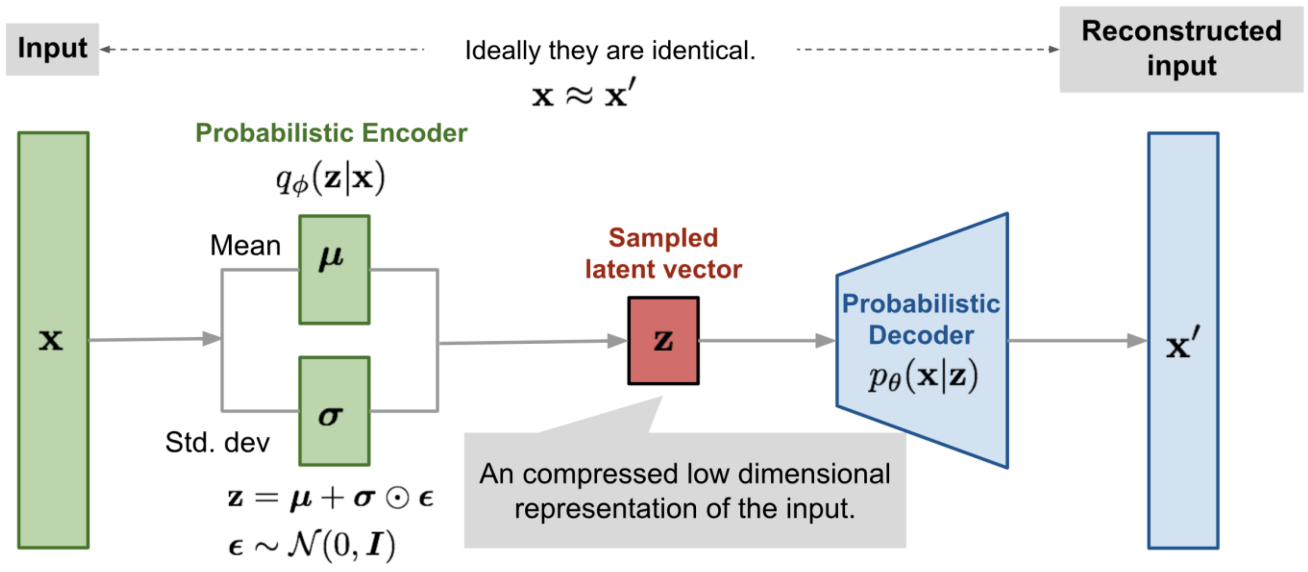
LSTM Architecture

- LSTM were designed to overcome the long-term dependencies problem in the RNN due to the vanishing gradient

How LSTM is actually doing it?

$$i_t = \sigma(x_t U^i + h_{t-1} W^i)$$
$$f_t = \sigma(x_t U^f + h_{t-1} W^f)$$
$$o_t = \sigma(x_t U^o + h_{t-1} W^o)$$
$$\tilde{C}_t = \tanh(x_t U^g + h_{t-1} W^g)$$
$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t)$$
$$h_t = \tanh(C_t) * o_t$$

- **Forget Gate**: The forget gate controls what information from the previous cell state should be discarded (forgotten).
- **Input Gate**: The input gate controls what new information from the current input should be added to the cell state.
- **Output Gate**: The output gate determines what part of the cell state should be output as the hidden state.



VAE(Variational Auto Encoder)

- A Variational Autoencoder (VAE) is a model that learns to encode data into a probabilistic latent space and then decodes it back, minimizing the difference between the original and reconstructed data to follow a known distribution (usually Gaussian)

Components of the VAE

- **Probabilistic Encoder**: In a Variational Autoencoder (VAE), the probabilistic encoder (also called the recognition model) maps the input data to a distribution in the latent space, typically a Gaussian, instead of a fixed point.
- **Sampled Latent Vector**: The latent vector is sampled from the distribution output by the probabilistic encoder.
- **Probabilistic Decoder**: The probabilistic decoder takes the sampled latent vector as input and reconstructs the original data. It models the likelihood of the data given the latent variable

Naïve Bayes

- Naïve Bayes was trained with the Kaggle data and then the accuracy were calculated on the synthetic data generated by the LSTM – VAE.
- The Gaussian Naïve Bayes gave the accuracy of 0.8.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

Results

LSTM-VAE-

- Reconstruction Loss = 0.876
- KL Divergence Loss = 0.32

Classification-

- Accuracy = 0.926
- Precision = 0.846
- F1-Score = 0.814