

Generative models

vs

Discriminative model

$$P(x, y)$$

Probability when x and y

Simultaneously happen

x is independent of y
and vice versa.

Probability that it will rain
if I carry an umbrella.

$P(x, y)$ = Joint probability,

no dep. var

Naïve Bayes Model

Transformer model

GPT

2017

Ashish Vaswani

Attention is what
you need

NLP

Where

the sky

☐

SVM Classify

RF regression

DT

LR

LR

$$\underline{y} = f(\underline{x})$$

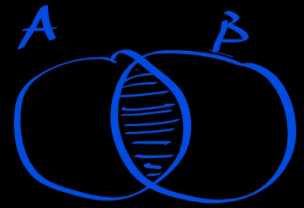
$$P(\underline{y} | \underline{x}) = \text{Conditional}$$

dependent var

NLP

modified neural network
that can predict next
token in a context

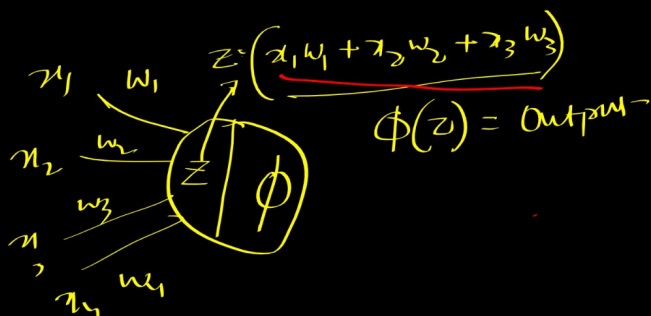
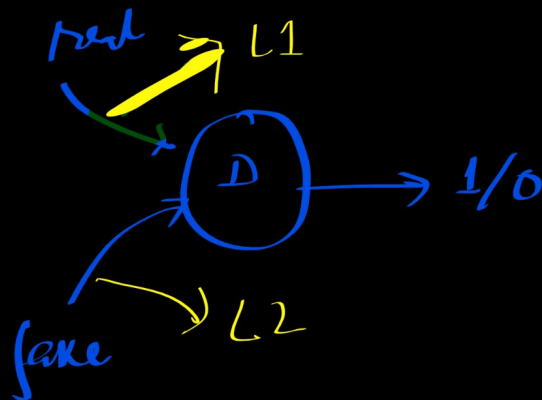
$$P(A|B) = \frac{P(A, B)}{P(B)}$$



$$P(A, B) = P(A|B) \times P(B)$$

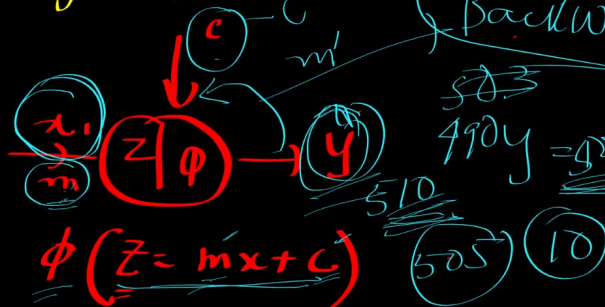
Generate
Models.

NN

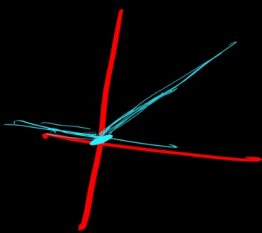
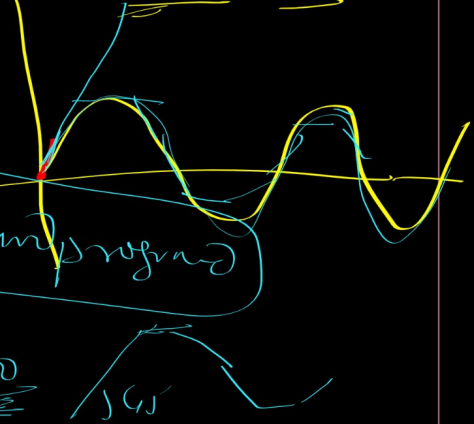


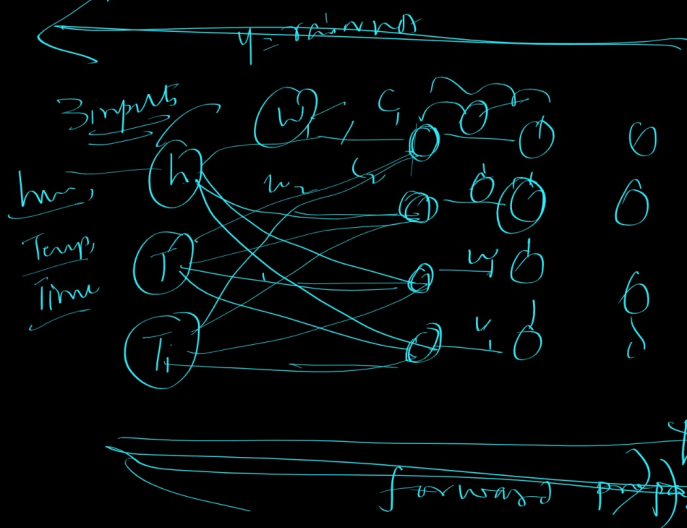
$$y = f(x)$$

Backward propagation



$$\sin(x) = f(x)$$





Loss func

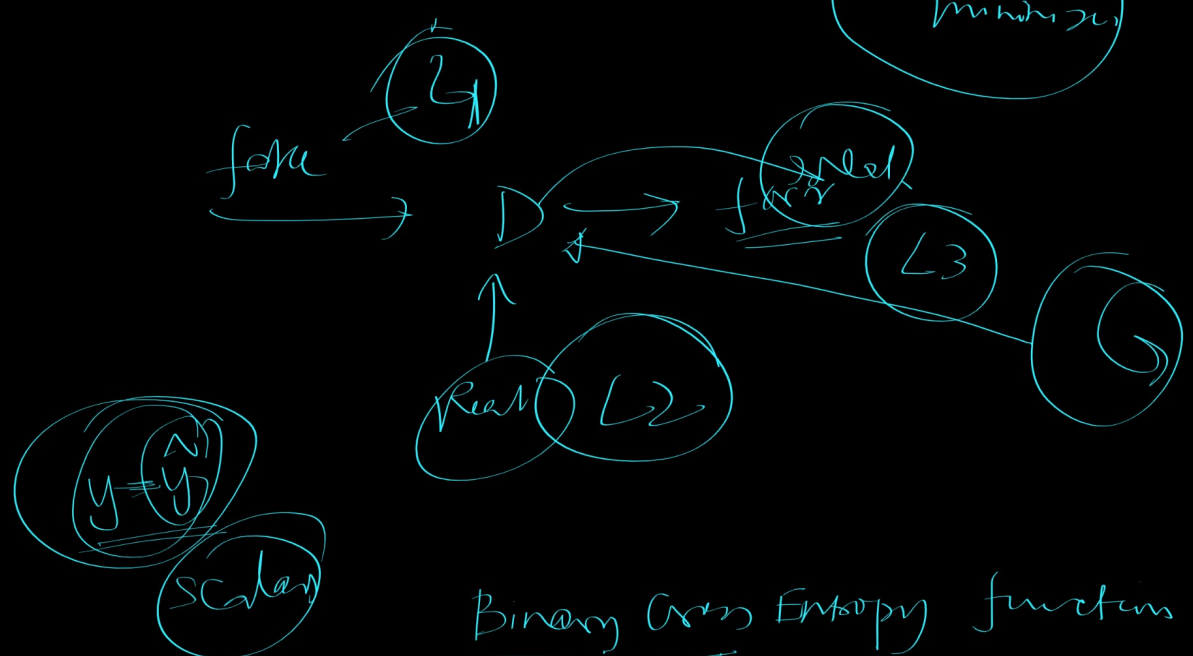
$$0.5 \ln(y - \hat{y})$$

0 → \hat{y}

h	T	T _i	y
21	20	10	1
			0

Loss()

minimize



Binary Cross Entropy function

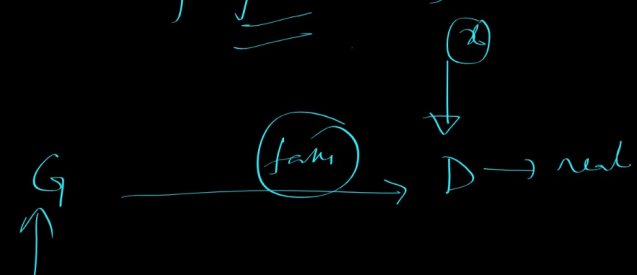
$\frac{\partial (L)}{\partial x} = 0$

$(L) = - \sum y \ln \hat{y} + (1-y) \ln(1-\hat{y})$

Gradient Descent = ln

for real imgs when $y = \underline{1}$, $\hat{y} = D(x) = L = \ln D(x)$

for fake imgs when $y = 0$, $\hat{y} = D(G(z)) = L =$



$\ln(D(G(z)))$

$-\ln(1-0)$

$\ln(1-D(G(z)))$

