T-SNE on MNIST

Group points based on their visual similarity.

Visual a 784d dataset $\xrightarrow{\text{to}}$ 2D
by embedding the dataset.

Entropy, Cross-Entropy and KL-Divergence

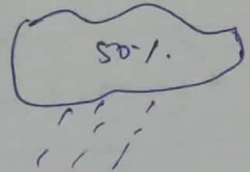
↓

Cost fn for classifiers

Claude Shannon → to transmit 1 bit of info is to reduce the recipients uncertainity by 2.

↓

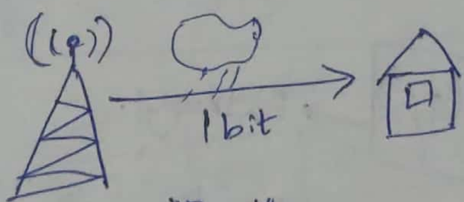
Information theory.



$P = 1/2$

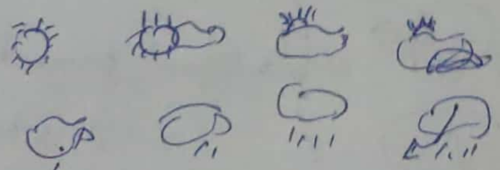
now $P = \frac{1}{2} \times 2 = 1$

$2^1 = 2$



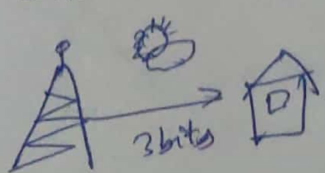
"Rainy"

40 bit message ⇒ but actually 1 bit of info.



⇒ Dividing your uncertainity by a factor of 8.

⇒ 3 bits of useful info.



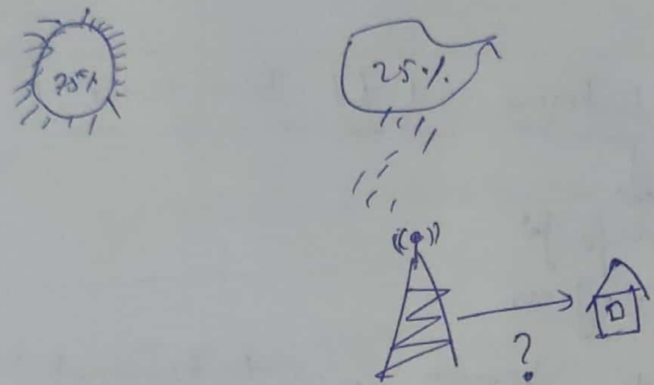
$2^3 = 8$

⇒ 3 bits.

$\log_2(8) = 3$

binary logarithm of the uncertainty reduction factor

$\log(1/x) = -\log(x)$



How much info. from the weather station are you getting on an average?

$$75\%\left[-\log_2(0.75)\right] + 25\%\left[-\log_2(0.25)\right]$$

$$= 75\%\,[0.41] + 25\%\,[2]$$

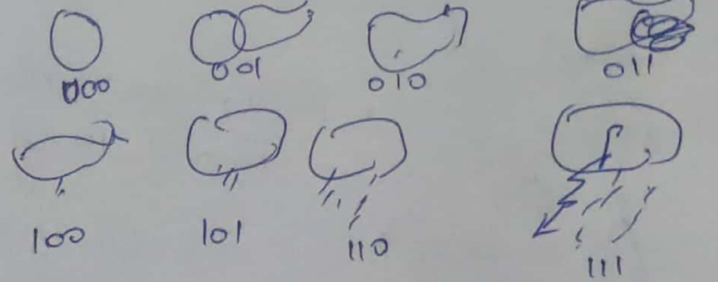
$$= \underline{\underline{0.81 \text{ bits.}}}$$

(or)

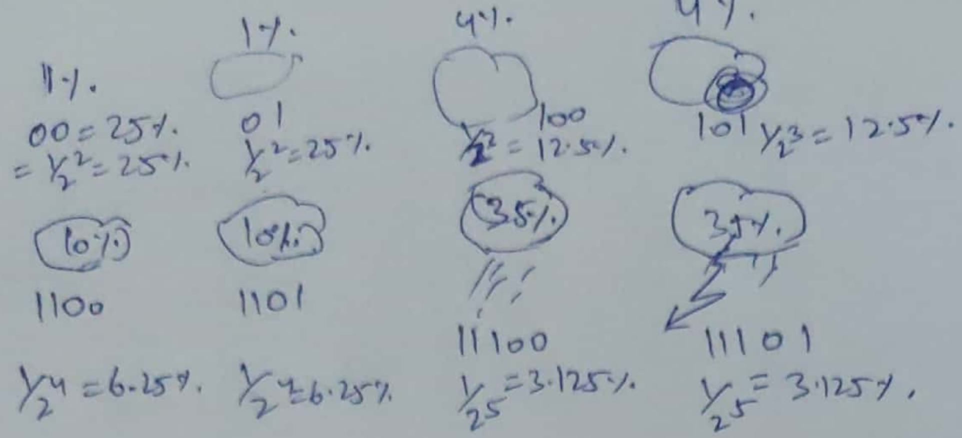
Avg. amt of info. from 1 sample from a probability distribution.

$$H(p) = \boxed{-\sum_i p_i \log_2(p_i)} \rightarrow \text{Entropy.}$$

Cross-Entropy $\rightarrow$ Avg. msg length. $\rightarrow$ 3 bits



$\Big\}\rightarrow$ Equal probability $\Rightarrow$ 3 bits



p = true distribution. (sums to 100%.)

q = predicted distribution. (Does not sum to 100%.)

$$\boxed{H(p) = -\sum_i p_i \log_2(p_i)} \longrightarrow \text{Entropy.}$$

$$\boxed{H(p,q) = -\sum_i p_i \log_2(q_i)} \longrightarrow \text{Cross Entropy.}$$

$$\boxed{\text{Cross Entropy} = \text{Entropy} + \text{KL Divergence}}$$

$$\boxed{\log_2(x) = \frac{\log(x)}{\log(2)}}$$

$$\boxed{D_{KL}(p\|q) = H(p,q) - H(p)}$$

→ 1-hot vector, i.e 1 class (true class) has a prob. of 100%.

	Cat	Dog	Fox	Cow	Panda	Bear	Dolphin
True Distribution	0%	0%	0%	0%	100%	0%	0%
Predicted Distribution	2%	30%	45%	0%	25%	5%	0%

$$\text{Cross Entropy loss} = -\sum_i p_i \log(q_i)$$

$$= -1 \cdot \log(0.25) = 1.386.$$

Incase this is 0.

$-1 \log(0) =$ undefined → cost will grow very large.

Incase this is 100%.

$-1 \log(1) = 0$

If both distributions are same $\boxed{C.E = \text{Entropy}}$