Probability

	R	G	B
Y : Actual probabilities	[0.35	0.25	0.4]
Y-hat: Predicted probabilities	[0.25	0.45	0.3]

Error In the prediction:

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \widehat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Can we do batter ??

With Certain Events

Certain event-> can still be written as Probability distribution
where whole mass located at one label

A	B	C	D
[0	1	0	0]
[0.2	0.1	0.6	0.1]

Classification Problem:
Probability distribution of true label possible

Expectation

- Expectation: For every event there is some value associated with it

A	B	C	D	
[0	1	0	0]	True Y
[0.2	0.1	0.6	0.1]	Predicted Y
[10k	20k	5k	2k]	Value[Rewards]

Expectation

- Expectation: For every event there is some value associated with it

$$
\begin{array}{lccccl}
 & A & B & C & D & \\
[& 0 & 1 & 0 & 0 &] \text{ True Y} \\
[& 0.2 & 0.1 & 0.6 & 0.1 &] \text{ Predicted Y} \\
[& 10k & 20k & 5k & 2k &] \text{ Value[Rewards]}
\end{array}
$$

$$\sum_{i=A,B,C,D} P(i)\, V(i)$$

Probability and Information Content

- High probability – Low Information Content
- Low probability – High Information Content
- A: Sun rises in east (Prob=1 means IC =?)
- B: there will be a cyclone tomorrow

$$I(x) = -\log P(x)$$

Entropy

- For every events we have IC[Info. content] (Value associated with that event)

$$H(X) = -\sum_{x} P(x)\cdot \log P(x)$$

$$= \sum_{x} \underbrace{P(x)}_{\text{The prob. of event X}} \cdot \underbrace{\log\left(\frac{1}{P(x)}\right)}_{\text{WHAT IS THIS ?}}$$

- If we have P(X) =1, the entropy is 0. It has 0 bits of uncertainty. (-log1 = 0)
- Note that “entropy” in information theory capture increasing randomness.
- **Cross Entropy??**