We want to find the joint probability distribution of $x, t$

finding $p(x,t)$ is very difficult to find, since we have limited data points

The moment we figure out $p(z)$, where $z$ is your data we can predict anything

Now if we, while finding $p(z)$, we find $q(z)$
How different is $q(z)$ from $p(z)$

We assumed initially that $t \sim N(y(x), \sigma^2)$
$p(x,t)$ could be assumed that way

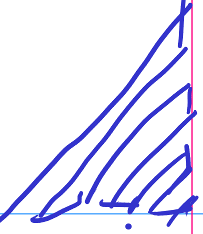
* When $x$ takes a very likely value, it won't have much info
* But if it contains a value unlikely to happen

① $\leftarrow$ $h(\cancel{x}=x) \propto \frac{1}{P(x=x)}$

② $\leftarrow$ $h(x=x \wedge x=y) = h(x=x) + h(x=y)$

independent events

③ $\leftarrow$ $p(x=x \wedge x=y) = p(x=x)\, p(x=y)$



$h(x) = -\log(p(x))$

$h(x,y) = -(\log(p(x)p(y)))$

$= -\log(p(x)) - \log(p(y))$

$= h(x)+h(y) \rightarrow$ 3 satisfied

$h(x) \propto \frac{1}{p(x)}$ isn't completely satisfied but kinda

$\boxed{h(x) = -\log(p(x))} \rightarrow \log\left(\frac{1}{p(x)}\right)$

uniform

| x | 1 | 0 |
|---|---|---|
| P(x) | 0.52 | 0.48 |

non-uniform

| x | 1 | 0 |
|---|---|---|
| p(x) | 0.1 | 0.9 |

$X = 1, 2, 3, \ldots, 100$

we need this to understand Decision Tree lol

Our average information is

$E(\text{Info}(x)) = -\sum p_i \log p_i$

which is Entropy